(1)

Contract No. N62269-90-D-0110
Report No. NAWCADWAR-93073-60

AD-A279 544

THE ADVANCED TECHNOLOGY CREW STATION: Development and Validation of a Workload Assessment Technique for Cockpit Function Allocation

Floyd Glenn, Ph.D., James Boardway and Robert Wherry, Jr., Ph.D.
CHI SYSTEMS, INC.
Gwynedd Plaza III
Bethlehem Pike At Sheble Lane
Spring House, PA 19477

David Cohen and Lt. Meghan Carmody
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

DTIC
ELECTE
MAY 23 1994
S B D

30 JULY 1993

FINAL REPORT

94-15370

Approved For Public Release; Distribution is Unlimited

Prepared for
Air Vehicle and Crew Systems Technology Department (Code 60B)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

94 5 20 184

NOTICES

REPORT NUMBERING SYSTEM — The numbering of technical project reports issued by the Naval Air Warfare Center, Aircraft Division, Warminster is arranged for specific identification purposes. Each number consists of the Center acronym, the calendar year in which the number was assigned, the sequence number of the report within the specific calendar year, and the official 2-digit correspondence code of the Functional Department responsible for the report. For example: Report No. NAWCADWAR-92001-60 indicates the first Center report for the year 1992 and prepared by the Air Vehicle and Crew Systems Technology Department. The numerical codes are as follows:

CODE	OFFICE OR DEPARTMENT
00	Commanding Officer, NAWCADWAR
01	Technical Director, NAWCADWAR
05	Computer Department
10	AntiSubmarine Warfare Systems Department
20	Tactical Air Systems Department
30	Warfare Systems Analysis Department
50	Mission Avionics Technology Department
60	Air Vehicle & Crew Systems Technology Department
70	Systems & Software Technology Department
80	Engineering Support Group
90	Test & Evaluation Group

PRODUCT ENDORSEMENT — The discussion or instructions concerning commercial products herein do not constitute an endorsement by the Government nor do they convey or imply the license or right to use such products.

Reviewed By: Joseph E. Kerr Date: 7 FEB 94

Contracting Officer's Technical Representative (COTR)
or
Point Of Contact (POC)

Reviewed By: John K Schmotl Date: 4 FEB 94

Branch Head

Reviewed By: T M Mitchell Date: 4 MAR 94

Division Head

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
30 July 1993

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
THE ADVANCED TECHNOLOGY CREW STATION:
Development and Validation of a Workload Assessment Technique for Cockpit Function Allocation

5. FUNDING NUMBERS
Contract No.
N62269-90-D-0110

6. AUTHOR(S)
Floyd Glenn, Ph.D. *
James Boardway *
Robert Wherry, Jr., Ph.D. *
David Cohen **
Lt. Meghan Carmody **

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CHI SYSTEMS, INC. *
Gwynedd Plaza III
Bethlehem Pike At Sheble Lane
Spring House, PA 19477

Air Vehicle and Crew Systems Technology Department **
(Code 6023)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCADWAR-93073-60

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Vehicle and Crew Systems Technology Department (Code 60B)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved For Public Release; Distribution is Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The design of a new severely upgraded aircraft cockpit requires many design decisions to be made, at least tentatively, prior to any opportunities for generation of detailed design specifications and experimentation with prototypes. In considering issues of interface design and function allocation, it is important to develop predictions concerning the effects of the various design alternatives on pilot performance. Task network models and workload estimation techniques are typically used jointly to accomplish this goal. The work described here was conducted in order to refine type of analysis and prediction technique as part of the U.S. Navy's Advanced Tactical Cockpit (ATC) Pilot-Vehicle Interface (PVI) Program.

14. SUBJECT TERMS
Advanced Tactical Cockpit (ATC), Pilot-Vehicle Interface (PVI), W/INDEX, Function Allocation Simulation System

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

TABLE OF CONTENTS

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

We have conducted research to develop and evaluate techniques for using projective workload assessment metrics in the process of determining appropriate function allocations for advanced tactical cockpits. We developed a task analysis of an air strike mission and used a network model as a framework for workload assessment. We then attempted a fairly straightforward W/INDEX-type estimation of workload with two experienced subjects. The results of that estimation process uncovered serious problems in both the very high correlations of the three cognitive channels that we used and also in the identification of workload thresholds after the application of the resource conflict components of the W/INDEX model. Factor analysis of the resource load estimates indicated only three or four independent factors, with no discrimination of separate resource factors within the cognitive channel. Accordingly, we have proceeded with attempts to develop a more suitable workload estimation framework that would solve these problems. At the same time, we have also sought empirical validation in the evaluation of the relative superiority of this new technique.

The workload assessment metric that we developed is based on concepts of time-constrained channel limits and time-based estimates of resource loads. For this second phase of workload evaluation, a revised technique was formulated in which five workload channels are defined (including a single cognitive channel) and loading on each channel is estimated in terms of the time demand for the resource on the channel rather than in terms of the effort demand. Task workload estimates were again made in independent fashion without regard to other activities that might or might not be concurrent with the task. Since the overall task time requirements and timeline were established via a mission analysis at the beginning of the study, subjects were asked only to estimate the proportion of time within each task that each resource would be used. The time measure produces a direct means for integration of demands across tasks, with a clear threshold of 100%.

In order to validate this time-based workload assessment technique, a software tool called the Function Allocation Simulation System was devised to step subjects through a tactical mission timeline, indicating all tasks which would have to be performed at each time when the mix of tasks would change. Subjects were asked to indicate which task would be selected for automation, based purely on

workload considerations, at each time step in the mission. We then compared the analytical workload judgements, generated by summing the percentage utilization of concurrent resources, with the task off-loading judgements to assess the consistency of the workload with the offloading assessments.

The implications and relations of this work to ongoing research in adaptive automation is discussed.

ACKNOWLEDGEMENTS

We wish to thank the subject matter experts in TACAIR operations who contributed extensively and very beneficially to this study. These include LCDR David Angrisani and Mr. Matt Sparta from NAWCADWAR and Mr. Peter Cotsonas from Veda Inc. We would also like to acknowledge the assistance of our colleagues at CHI Systems in both the conduct of the research and documentation of these results.

BACKGROUND

The design of a new or severely upgraded aircraft cockpit requires many design decisions to be made, at least tentatively, prior to any opportunities for generation of detailed design specifications and experimentation with prototypes. In considering issues of interface design and function allocation, it is important to develop predictions concerning the effects of the various design alternatives on pilot performance. Task network models and workload estimation techniques are typically used jointly to accomplish this goal. The work described here was conducted in order to refine this type of analysis and prediction technique as part of the U.S. Navy's Advanced Tactical Cockpit (ATC) Pilot-Vehicle Interface (PVI) program.

For the purpose of evaluating workload in a prospective cockpit design, we are interested in prospective (or projective) workload estimation techniques. Thus, we must focus on the subjective estimation of workload based on analyses of the tasks to be performed by the pilot. In proceeding with this evaluation, there are two major issues to be addressed: how to decompose the workload representation and how much to decompose the pilot's tasks.

Although the earliest representations of workload postulated a monolithic workload construct, with workload being represented as a single undifferentiated quantity, the concept soon developed that workload might more appropriately be treated as a multidimensional construct. In the multidimensional case, total workload can be defined as some functional combination of component load values. Several different approaches to the identification of workload dimensions have been employed, offering a variety of associated benefits for the analyses of designs and performance. Some techniques designed for retrospective assessment have focussed on affective aspects of the workload experience. Two well-known examples include the Subjective Workload Assessment Technique (SWAT), which discriminates the dimensions of time load, mental effort load, and psychological stress load (Reid, Shingledecker, & Eggemeier, 1981), and the Task Load Index (TLX) which includes the six dimensions of frustration, effort, performance, temporal demand, physical demand, and mental demand (Hart & Staveland, 1988). Although these techniques have been adapted for prospective assessments, the affective dimensions (e.g., psychological stress or frustration) are difficult to address in the prospective mode. Hence, alternative techniques that

employ dimensions associated with performance resources have generally been preferred.

The concept of performance resources divides human information processing into several distinct channels: capabilities for processing sensory inputs, for internal cognitive processing of information, and for effecting output actions for control of systems and operation of user interfaces. A fairly simple, widely-used technique embodying this concept is the McCracken-Aldrich technique, which defines workload in terms of the dimensions of visual, auditory, cognitive, and psychomotor resource channels (McCracken & Aldrich, 1984). This technique has been incorporated into the automated workload analysis tools TAWL (Bierbaum, Fulford, & Hamilton, 1989) and MAN-SEVAL (Laughery et al., 1988). Within these techniques, the load which a task imposes on each of the resource channels is estimated on a seven point scale. The total workload is then determined by adding the loads across the four channels and also across all tasks that are performed simultaneously at each point in the task timeline. It is assumed that there is some critical threshold such that performance will degrade or disintegrate when higher workload values are experienced; the natural candidate for such a threshold would seem to be 7 since it is the maximum value for each of the individual scales, but other values have also been used.

The McCracken-Aldrich technique is fairly easy to interpret and apply, but it has been criticized both for distinguishing too few resource channels and for aggregating individual channel loads into overall workload via too simplistic of an additive model. Wickens (1984) has argued for a representation of workload which incorporates the concept of possible conflicts between resource channels, with some channels exhibiting high conflicts with one another (e.g., the conflict of a channel with itself when it is to be used simultaneously on different tasks) and with other channels having relatively low conflicts (e.g., visual and auditory input processing channels). Wickens concept, known as Multiple Resource Theory, has been formalized in a workload aggregation formula and an automated workload analysis tool known as W/INDEX (North & Riley, 1988). The aggregation formula postulates that there is a conflict parameter which applies to every pair of resource channels (including each channel with itself) and which determines the proportional increase in workload when the channels must be used simultaneously rather than separately. W/INDEX is designed to allow arbitrary definition of the number and type of resource channels that contribute to workload. Much of the use by the tool developers is based on the assignment of specific interface display and

control components as resource channels. However, the W/INDEX developers have also identified general resource channels and associated values for channel conflicts. The most complete set of general resource channels that they have suggested includes the discrimination of three cognitive processing channels (spatial, verbal, and analytical) in addition to the two input channels (visual and auditory) and the two output channels (manual and speech).

Somewhat orthogonal to the issue of resource decomposition are the issues of time and task decomposition. All of the subjective workload assessment techniques discussed so far can be applied with an arbitrarily fine or coarse resolution of tasks and time. Workload assessments can be made at the highest level task, covering the entire time frame of performance with a single estimate (or group of estimates for multiple resources), or at intermediate levels of resolution down to very detailed perceptual, cognitive, and motor actions. It seems to be generally assumed that greater task decomposition will lead to greater fidelity of workload estimation, but there is very little empirical basis for this assumption. One relevant evaluation pertaining to this issue was reported by Card, Moran, and Newell (1983) with regard to time estimates for a text editing task.

The remainder of this paper briefly recounts two alternative methodologies that were successively developed for prediction of workload in a task network context. In the first, we employed a variant of the W/INDEX technique and investigated issues associated with the interpretability of the results and the general quality of the data obtained. In the second part of the study, we developed and evaluated an alternative, time-based technique for workload estimation and examined its validity by comparing workload profiles over time with separate decisions of task shedding made while reviewing a mission scenario timeline.

PART 1 - BASELINE WORKLOAD ESTIMATION

In the first part of our study, we attempted to conduct task and workload analyses using existing tools and techniques. We developed a task network simulation for an air strike mission (though Combat Air Patrol and Deck Launched Interceptor missions were also analyzed as part of this effort). Each of ten phases of the strike mission were implemented as task network models using MicroSAINT. The task network models were constructed in a completely deterministic form in order to conform precisely to a pre-established mission timeline. Thus, the task network models were employed primarily as a vehicle for computation of the workload function, rather than their more typical use for timeline generation. The workload measure used in the study was the W/INDEX model (North & Riley, 1988), which calculates workload as the sum of the loading on each of seven distinct channels plus penalties for between and within channel conflicts. In the task network simulation, all tasks were assigned workload values for each of the seven channels.

Subject Matter Experts

Resource effort estimates were provided by two recently retired U.S. Marine Corps pilots (P1 and P2 individually). Both of these pilots had significant operational experience (approximately 1000 hours) in the F/A-18 Hornet, which is an antecedent to the next-generation fighter/attack aircraft, as well as combat experience in the F-4 Phantom II. In addition, both pilots had assisted in the development of the strike mission scenario and the stipulation of the aircraft capabilities and, therefore, were intimately familiar with the tasks that were rated.

Workload Estimation

The pilots were asked to rate the amount of effort that would be required in each of seven human resource channels in order to perform each of 225 strike tasks. These channels included: visual perception, auditory perception, spatial information processing, analytical information processing, verbal information processing, manual activity, and speech. An eight point scale was used in which "0" indicated "no effort required" and "7" indicated "maximum effort required." They were also requested to estimate the overall effort needed to complete the task without the partitioning of resources. The pilots were instructed to rate each task

and/or each component of a task independently of any concurrent task or component. These estimates were gathered and recorded using a HyperCard program running on a Macintosh SE computer. Figure 1 shows the display interface used for data collection. Details on the definition of the resource categories, the data collection procedures and the construction of the data collection system can be found in Glenn, Cohen, Barba, and Santarelli (1990).

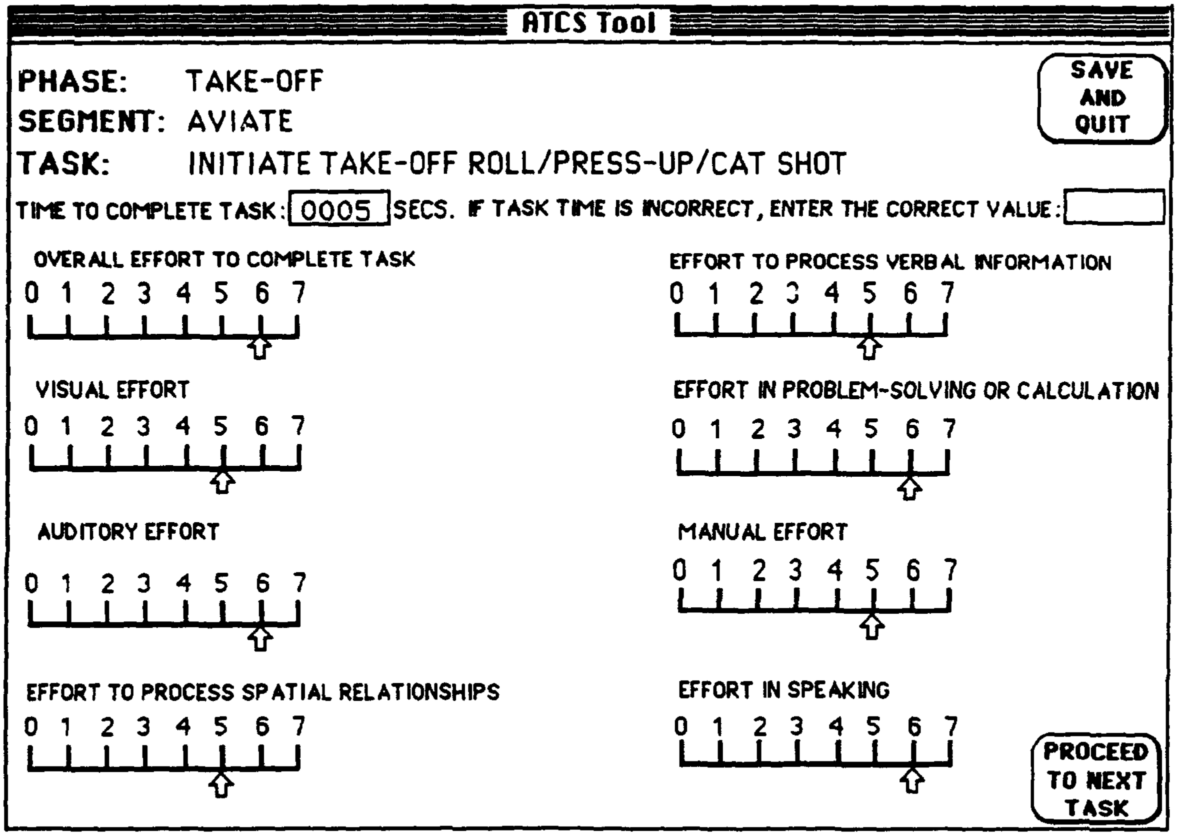


Figure 1 -- Display Screen for Data Collection for Part 1 Study

Network Simulation Construction

MicroSaint simulation software running on a 386 personal computer was used to implement task network representations of the strike mission. MicroSaint, a product of Micro Analysis and Design Inc., allows the user to develop, execute, and analyze the results of network simulation models. Models are constructed by defining task nodes and connecting them together via branching or control logic to form a task network. A task node consists of its associated attributes, which usually includes: task identification, mean execution time, beginning and ending effects, and following task information. When the simulation is executed, the software

provides the ability to capture data on the state of the simulation. For a more comprehensive description of MicroSaint and its application to a tactical mission (for the LHX helicopter) see Laughery, Drews, and Archer (1986).

The required models were constructed for each of the ten phases of the strike mission: take-off, climb, cruise out, descent, ingress, attack, egress, climb (second), return to force, and recovery. The timeline for each phase was further decomposed into segments within mission phases (e.g., aviate, navigate, etc.) and individual tasks (e.g., monitor system status) using the task analyses as a reference (Cohen, 1990). The models were developed from an analysis of the strike mission timelines (Veda, 1990). Task networks were then created by assigning connections between tasks on the basis of task execution times and logical heuristics. Task start times and durations were acquired from the timelines and later verified by subject matter experts. Mission segments were used as the starting point for all tasks within that segment. In the models, mission segments can be considered pseudo-tasks because although they have no time or workload charges associated with them, they were needed to provide the grouping for tasks. Figure 2 shows an example of the network diagrams that were drawn to represent the structure of the task relationships (see Glenn et al., 1990).

After the task network diagrams were developed, they were implemented in MicroSAINT. Network models were built using the task connections shown in the network diagrams and the task timing information obtained from the timelines. The release condition for each task contains a function (i.e., logical and mathematical control statement) which forces the task to execute at the correct time to effectively mimic the timeline. Mean execution times for tasks were taken directly from the timelines. When tasks repeated more than once with different task durations, a variable was inserted as the mean time. Functions were written to insert the correct ti· e value into the mean time variable at the appropriate time. Task beginning effects contained the workload values across the seven channels (described below) for all the tasks. When a task was executed, its associated workload values became active which caused them to be included in the workload calculation. Task ending effects contained zeros for all channels to initialize the task workload values. Tasks which could follow execution of some other task were assigned on the basis of the examination of the timelines. The probability of taking a following task (which was always set to 0 or 1 via program logic) contained functions which controlled branching to other tasks or back to itself, if that task was iterative.

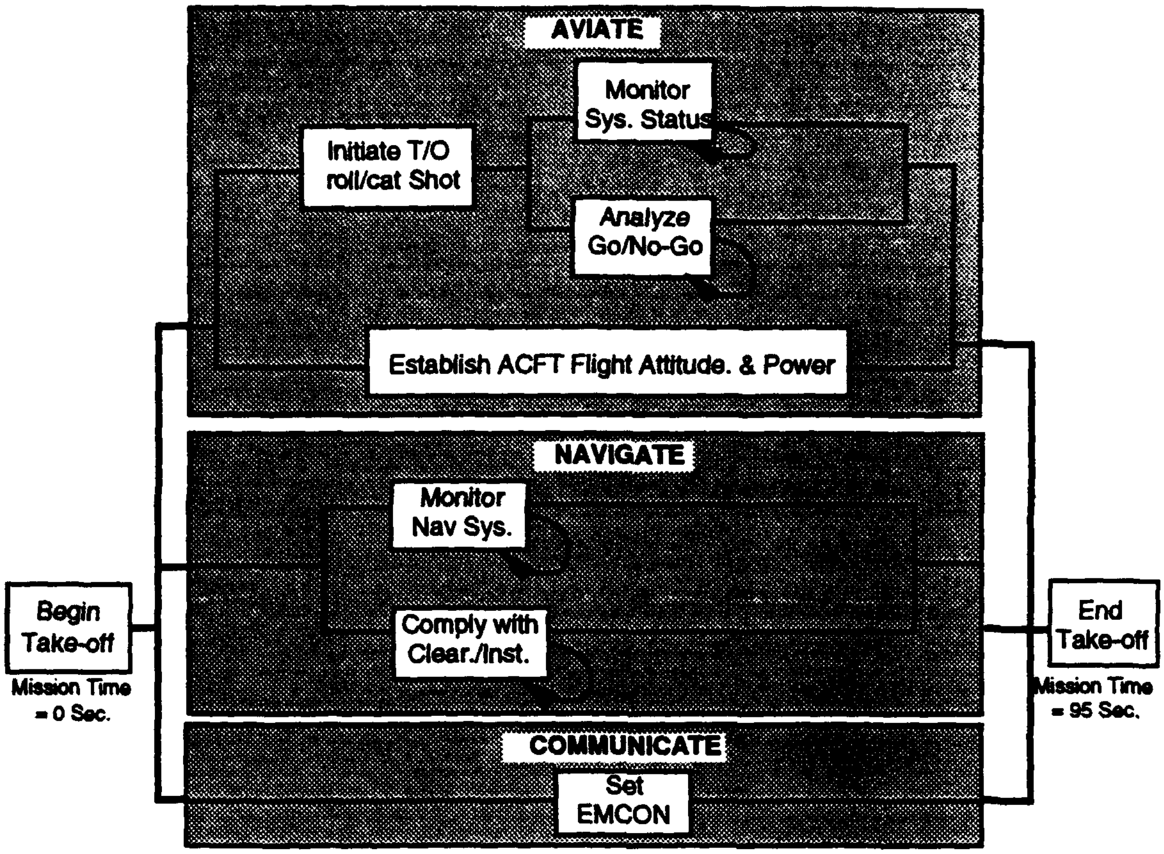


Figure 2 -- Portion of Task Network Diagram for Strike Mission

The simulations were set to use a one second time step so that workload would be calculated for each second. In addition to workload (which is defined as the total loading according to the W/INDEX equation), individual channel loading values were also captured at one second intervals. The simulations which were created in this effort were both fully deterministic and clock-driven. The simulations will yield the same results each time they are run and these results are tied directly to the clock. This was done to ensure that all tasks begin and end at the correct time and conform to the pre-established strike timeline.

Workload Model

The function to calculate workload based on the subjective ratings was the instantiation of the W/INDEX algorithm. Total workload was divided into components based on the subjects' estimates of the effort taxing the seven resources. The first two channels (visual and auditory) represent input channels. The next three channels (spatial, analytical, and verbal) represent cognitive

processing channels. The last two channels (manual and speech) correspond to output channels. Within each task network, all tasks were assigned workload values for each of the seven channels. These values were valid for the duration of the task.

The W/INDEX algorithm used these estimates to calculate workload according to the following expression:

$$W_T = \sum_{i=1}^{I} \sum_{t=1}^{m} a_{t,i} + \sum_{i=1}^{I} \left[(n_{t,i} - 1) c_{ii} \sum_{t=1}^{m} a_{t,i} \right] + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} c_{ij} \sum_{t=1}^{m} (a_{t,i} + a_{t,j})$$

where:

W_T	=	instantaneous workload at time T
i,j	=	1...I are the resource channels
t	=	1...m are the tasks occurring at time T
$n_{t,i}$	=	number of tasks occurring at time t with nonzero load values for channel i
$a_{t,i}$	=	load value for channel i in performing task t
$a_{t,j}$	=	load value for channel j in performing task t
$c_{i,j}$	=	*conflict* between channels i and j
c_{ii}	=	conflict within channel i

(NOTE: The third term of the W/INDEX algorithm is only calculated when both $a_{t,i}$ and $a_{t,j}$ are non-zero.)

Note that the three additive terms in the above formula correspond respectively to raw workload (i.e., the simple sum of resource loads across tasks), within-channel conflicts (i.e., conflicts arising from simultaneous use of the same channel on different tasks, and between-channel conflicts (i.e., conflicts between different channels on different tasks). One of the major features of the W/INDEX algorithm is its use of a conflict matrix to assess the workload penalties associated with these between and within channel conflicts. The conflict matrix that was used in these simulations consists of 28 terms which represent the conflict of each of the seven channels with itself and all other channels. The conflict coefficients (Figure 3) were adapted from the research of North and Riley (1988) and ranged from 0 to 1. A technical discussion of the implementation of the features of multiple resource theory into the task network simulation (including the function source code) can be found in Glenn et al. (1990).

		Input		Internal Processing			Output	
		Visual	Auditory	Verbal	Spatial	Analytical	Manual	Speech
Input	Visual	0.8	0.3	0.2	0.3	0.2	0.2	0.2
	Auditory		0.8	0.3	0.2	0.2	0.2	0.5
Internal Processing	Verbal			0.2	0.2	0.2	0.2	0.3
	Spatial				0.3	0.3	0.3	0.2
	Analytical					0.1	0.2	0.2
Output	Manual						0.8	0.[illegible]
	Speech							1

Figure 3 -- Conflict Coefficients for W/INDEX Model used in Part 1

PART 1 - RESULTS

Correlations and Factor Analysis

Means, standard deviations, and correlations of the workload ratings of the seven resources across all tasks were obtained independently for both P1 and P2. Relatively high intercorrelations among all seven resource channels and extremely high correlations among some of them suggested that raters must have felt that many tasks required all of the "independent" resource channels or that the raters were unable to discriminate among them. At the very least, the raters appeared to be indicating that whenever high effort levels were required by any input resource channel, high effort levels would also be required for cognitive and output channels as well. To identify the number and nature of independent factors causing the high intercorrelations among the seven postulated resource channels, Principal-Axis (PA) factor analyses of the intercorrelations for each subject were accomplished. For these analyses, initial communalities (h^2s) for each factor analysis were estimated using the highest-r method. Solutions were iterated until beginning and ending communality estimates stabilized within .001. Four factors were extracted for each pilot. Varimax-rotated factors failed to yield simple structure (i.e., where some variables have high loadings on a factor and all others have zero loadings) for the factors for either pilot. Ultimately, graphical rotation was used to identify the general factor responsible for the extremely high intercorrelations among the seven resource channels. Results of those analyses are shown in Table 1.

*Table 1 -- Correlations and Factor Loadings for Part 1 Study

Data analysis results for pilot 1 (P1):

resource channel	Mean	S.D.	Correlations 1	2	3	4	5	6	7	factor loadings 1	2	3	4
1 visual	3.13	2.05		585*	954	673	924	928	779	**973**	008	041	-031
2 auditory	1.21	1.40			570	805	628	566	628	**597**	**689**	007	001
3 spatial	2.99	2.23				653	930	908	764	**981**	-024	-045	-036
4 verbal	1.26	1.39					727	657	835	**693**	**568**	013	**349**
5 analytical	2.87	1.98						891	811	**951**	086	-008	057
6 manual	2.72	2.01							818	**940**	004	**295**	002
7 speech	1.36	1.76								**807**	**210**	**193**	**438**

Data analysis results for pilot 2 (P2):

resource channel	Mean	S.D.	Correlations 1	2	3	4	5	6	7	factor loadings 1	2	3	4
1 visual	2.90	1.63		448	788	578	515	559	276	**982**	002	027	019
2 auditory	1.38	1.30			391	517	393	320	539	**453**	**579**	124	012
3 spatial	2.94	1.95				497	650	548	309	**787**	-002	177	**395**
4 verbal	1.97	1.41					503	323	400	**582**	**434**	000	127
5 analytical	2.50	1.50						281	285	**513**	273	003	**620**
6 manual	2.03	1.73							556	**553**	-010	**646**	-003
7 speech	.93	1.23								**261**	**584**	**644**	-020

*three decimals omitted for values other than means and standard deviations and variance portion

The sum of the eigenvalues (i.e., the sum of the *resource channels' variance* explained by each factor) and the sum of the communalities (i.e., the sum of each variable's variance explained by all of the factors) show that 92.6% of the variance of all variables across all tasks was explained by P1's four factors. For P2, the comparable figure was 73.4%.

Interpretation of the Rotated Factors

Both pilots yielded a very strong general factor (i.e., one in which all variables have high loadings) that loaded most highly (.973 and .982, respectively) with the visual input channel. The second highest loadings on those factors was the spatial information processing channel (.981 and .787). This indicates that both pilots perceived that when the tasks being rated were dominated by visual inputs, they also required spatial processing. Because all of the other channels loaded significantly on this **visual-spatial factor** (factor 1), it indicates that the tasks dominated by visual-spatial demands were sufficiently complex to demand the other resource channels as well (e.g., analytical thought, verbal communications, and manual outputs).

A second and independent **verbal-communications factor** (factor 2) was also found for both pilots. It was dominated by high loadings on auditory input, verbal information processing, and speech output. This factor indicates that the pilots also distinguished tasks that were dominated by (or required relatively more or less) verbal communications.

A third and independent **manual and speech output factor** (factor 3) was also found for both pilots, although with somewhat weaker loadings for P1. This factor indicates that the pilots distinguished among tasks that required relatively more or less output demands.

While an additional independent factor was found for each pilot (factor 4), the nature of their final factors appeared to be quite different. For P1, the final factor loaded highest on verbal information processing (.349) and speech output (.438) indicating P1 differentiated among tasks that required more or less **speech production** than would have been indicated by the loadings for the resources on the visual-spatial or verbal-communications factors. For P2, the final factor had high loadings on the analytical (.620) and spatial (.395) information processing channels, indicating that P2 may have made finer distinctions concerning the amount of **analytical thought required for spatial tasks**.

By far the most variance of the ratings for both pilots was explained by the first factor. This suggests that differential workload ratings (at least for these tasks) were determined primarily on the basis of the extent to which the visual-spatial factor was important to the rated tasks.

Workload Predictions

Because of the close agreement of the workload ratings provided by the two subjects, summary workload predictions are presented based on the average ratings of these subjects. Figure 4 presents the profile of total instantaneous workloads calculated with the W/INDEX model defined above. The workload values clearly vary widely, both from moment to moment and across the various phases of the mission. During the Cruise phase, for example, the workload values vary from 30 to 170, with an average of about 75 for the phase. During the Attack phase, on the other hand, the values range from 200 to 5500, with an average of 1150. Note that these values are still based on the original effort scale of 0 to 7 on which subjects made component resource estimates.

The contributions of each of the separate resource channels to the overall workload profile are illustrated in Figure 5. The values in the figure represent averages across each mission phase in order to facilitate summary comparisons. These values include the relevant within-channel conflict terms from the above workload formula. Note that the visual and manual resources seem to dominate across all phases, though especially in the highest-workload attack phase.

Relative *contributions* of the within-channel conflict and between-channel conflict terms in the workload formula are presented in Figure 6. For the three mission phases with the highest overall workload (i.e., Ingress, Attack, and Egress), the majority of total workload is generated by the within-channel component and the second greatest contributor is between-channel conflict. It is interesting to note that, for these three mission phases, the raw workload component (i.e., the simple sum of channel load values) accounts for less than 20% of the total workload, and considerably less than either of the conflict terms. This pattern is considerably different in the case of the other seven mission phases where the total workload is much lower and the three components (raw workload and the two conflict terms) all are roughly comparable in magnitude.

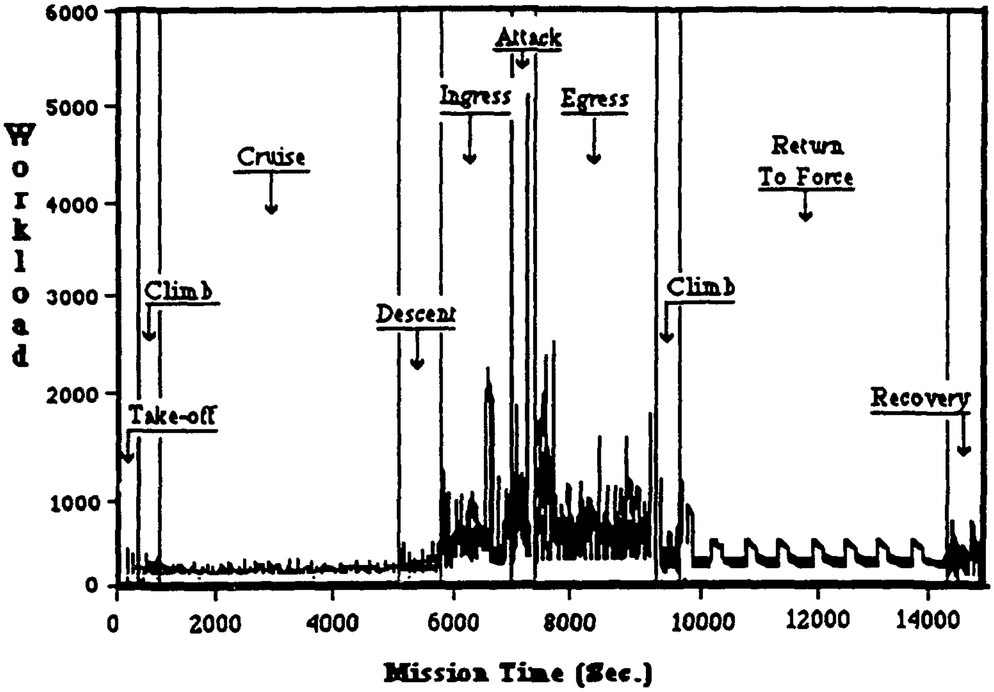


Figure 4 -- Profile of Total Instantaneous Workloads over Mission Timeline (Part 1)

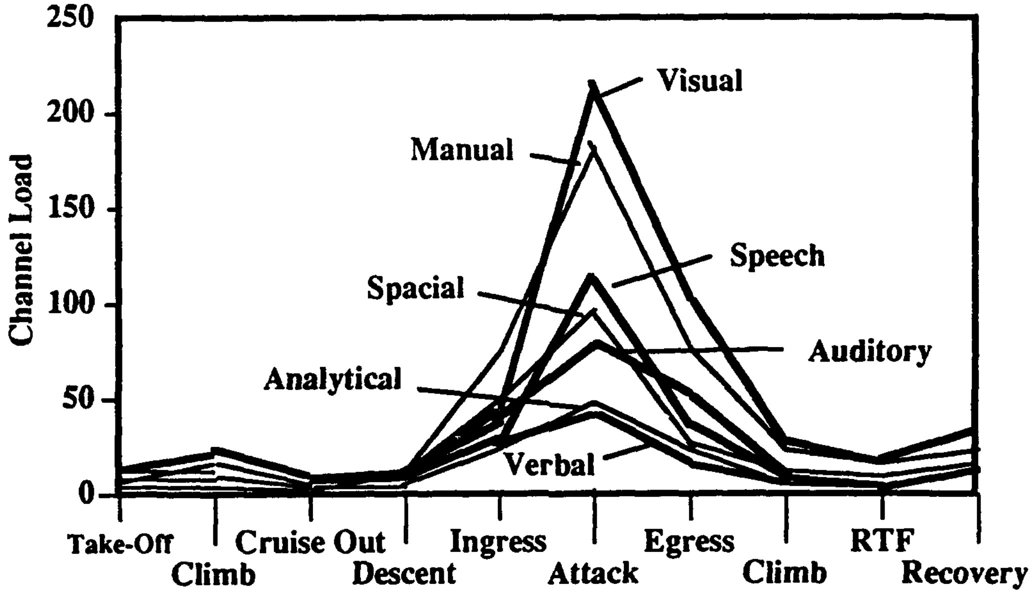


Figure 5 -- Contributions of Each Resource Channel to Overall Workload

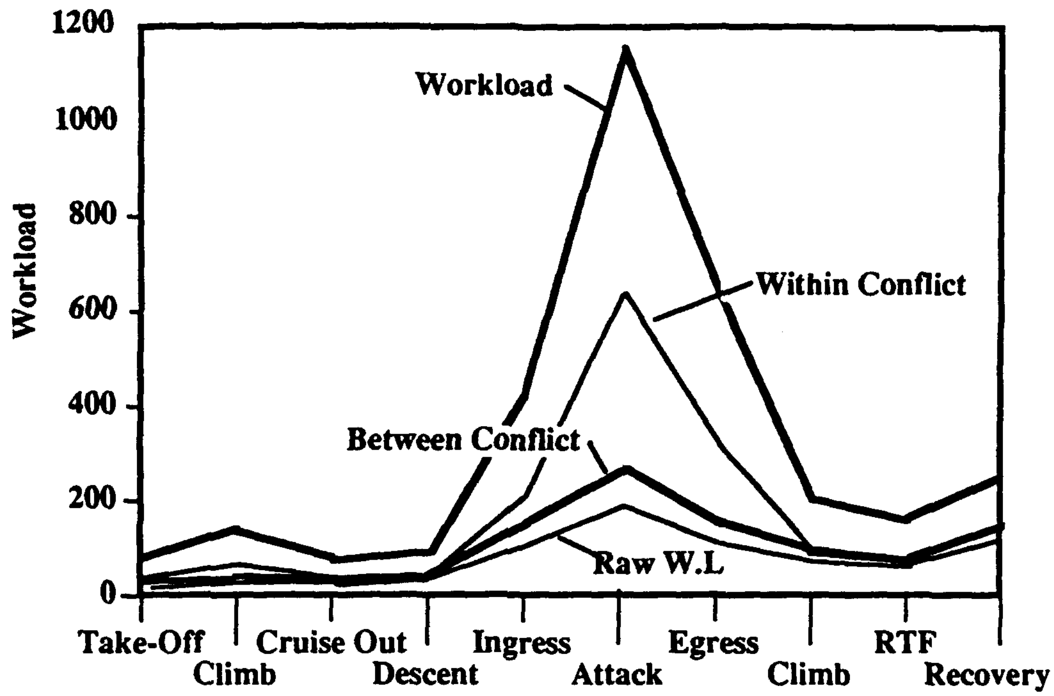


Figure 6 -- Contributions of Conflict Terms and Raw Workload

DISCUSSION OF PART 1 - BASELINE WORKLOAD ESTIMATION

Although seven independent channels were postulated, it is clear that their rated usages were highly related for the tasks studied. Three common independent factors emerged across the subjects: a visual-spatial factor, a verbal-communications factor, and a manual-speech output factor. This strongly suggests that the seven channels are highly confounded in real-world tasks. Consequently, subjects cannot make independent estimates of these resources. This is especially evident since most of the variance of the ratings was explained by the visual-spatial factor and therefore the difference in workload ratings across tasks was determined primarily by the extent to which this factor impacted the rated tasks.

A series of Pearson correlations established the real productivity of the approach used. For example, correlations between the total raw workload (addition of the seven channel estimates for all active tasks), the W/INDEX calculated workload (which includes within and between channel conflicts), the total overall workload (addition of the estimates of overall workload) and the number of active tasks for the phases of the strike mission yielded no $\underline{r}$ below 0.9. Essentially, use of a conflict matrix and segregating effort into the seven channels did not produce a predictive power superior to the number of tasks alone. These results are in accord with the results of the factor analysis – subjects' estimates were heavily influenced by a single "overall" factor with a magnitude related to the number of active tasks.

Other concerns also became evident. Most workload techniques employ a rating scale that does not have a well founded threshold. In our approach, like many others, we used a 0 to 7 scale. Unfortunately, simply adding up the estimates across active tasks generates large workload values which are not meaningful. Then, including the conflict matrix increases the values even further. This leads to a serious problem in identifying a threshold for workload with predictable performance consequences if that value is exceeded.

These three problems (discrimination of resource channels by subjects, calibration of the workload scale, and identification of a threshold) provided the motivation for a second phase of this study in which we sought to develop an alternative workload function and a means for its validation.

PART 2 - TIME-BASED WORKLOAD AND VALIDATION

The second part of this study addresses methodological problems encountered in the previous effort, including: reliability of subjective estimation, lack of thresholds, channel independence, and validity of the W/INDEX channel conflict matrix. In order to determine whether subjects can provide reliable subjective estimates of workload for separate resource channels, the estimation procedure was simplified and the data collection procedure was also modified.

In order to develop an alternative workload function which would overcome the problems identified with the baseline workload concept, we examined prior analyses of workload that we had conducted using a simulation tool known as the Human Operator Simulator, or HOS (Lane, et al., 1977, 1981). The HOS workload concept is that the human operator can perform multiple simultaneous tasks by switching attention rapidly back and forth between tasks, with performance resources (i.e., cognition, vision, hands, etc.) constrained to perform one action at a time but with multiple resources capable of operating in parallel. The limit on workload is reached whenever any resource is unavailable to perform required functions. Workload analyses were conducted with HOS simply by using the simulation to generate timelines of predicted performance and then comparing aspects of the timeline to required performance milestones and features; failures to satisfy requirements were interpreted as indications of excessive workload.

In order to convert this time-based workload representation from the simulation domain to the domain of subjective-prospective workload estimation, we sought to ask our expert subjects to make the same kind of resource utilization predictions that we had obtained from HOS -- How much is each resource being used by each task during each time interval? Since we have already established an application context in which task times have been firmly defined by separate mission and task analyses, we chose simply to ask the subjects to estimate the percentage of time that each resource channel would be used for each task. As before, these estimates were obtained by focusing on each task in isolation from all other tasks and the mission timeline. As a simplifying assumption, we treated all estimates of resource utilization as occurring homogeneously during the course of the task performance period. For example, if a task was estimated to last for 10 seconds and the visual channel was estimated to be required for 25% of the task duration, then we simply assumed that the visual channel was used for 25% of

every second (or other smaller or larger interval of analysis) over the course of the 10 seconds of task performance. By homogenizing the resource estimates in this fashion, we can then aggregate across tasks being performed simultaneously and determine, on a moment by moment basis, the total percentage utilization of each resource channel. Our expectation is that the human will be overloaded whenever the total utilization on any resource channel exceeds 100%. However, brief episodes with small excesses are not expected to be of any consequence because of the capability of the human to employ dynamic rescheduling strategies to make non-homogeneous use of resources and so avoid the overloads.

In order to resolve the problem which the baseline workload technique encountered with regard to discrimination of resource channels, we reduced the total number of channels from seven to five for this revised technique. We collapsed the three cognitive channels that we used in the baseline technique (i.e., spatial, analytic, and verbal) into a single cognitive channel and retained the other four channels as defined in the baseline (i.e., with visual and auditory input channels and manual and speech output channels).

The specific predictions of overload points that are provided by this revised workload representation create a clear opportunity for validation of the technique. After we have obtained the resource load estimates for individual tasks and generated the timeline of resource loadings for the mission scenario, we can ask the subjects to review the timeline of tasks and indicate which, if any, should be off-loaded in order to maintain a manageable workload. Agreement between the analytic predictions and the offloading judgements would constitute a type of validation for this workload estimation scheme.

This part of the study examined only one phase of one mission scenario -- the attack phase of the strike mission, because this was found to be the highest workload phase in the baseline study.

Subject Matter Experts

Resource effort estimates and task shedding judgements were provided by three recently retired pilots. Each of the subjects had significant operational experience (approximately 2000 hours) in the F/A-18, F-4, and training aircraft. Two of the subjects were U.S. Navy pilots whose primary experience was in the F/A-18. The third subject had a similar amount of experience as a Weapons Systems Officer for the U.S. Air Force in the F-4E aircraft. Unlike the subjects in the

first part of this study, these three subjects had not been involved in the earlier development and analysis of the strike mission scenario.

Workload Estimation

The pilots were asked to rate the amount of resource utilization required in each of five human resources or channels in order to perform each of 40 tasks involved in the attack phase of the strike mission, using the same scenario as in the baseline study. These channels include: visual perception, auditory perception, cognitive processing, manual activity, and speech. A percentage scale was established as the basis for resource utilization estimates and the subjects were required to specify their estimates using just five points on this scale: only the values of 0%, 25%, 50%, 75%, and 100% were allowed. We arrived at this scale based on the observation that subjects would probably only be able to distinguish about 7±2 points on the percentage continuum, and this five point scale seemed particularly familiar. The pilots were instructed to rate each task and/or each component of a task independent of any concurrent task or component. These estimates were gathered and recorded using a paper form, a sample of which is illustrated in Figure 7.

TASK:
COMPARE PRESENT STATUS TO MISSION PLAN
Duration = 5 sec.
Corrected Duration = __________ sec

	Seeing	Hearing	Thinking	Manual	Speaking
% Time*					

TASK:
COMPLY WITH CLEARANCE/INSTRUCTION
Duration = 4 sec.
Corrected Duration = __________ sec

	Visual	Auditory	Cognitive	Manual	Speech
% Time*					

TASK:
COMPUTE TIME ON TARGET (TOT)
Duration = 5 sec.
Corrected Duration = __________ sec

	Visual	Auditory	Cognitive	Manual	Speech
% Time*					

TASK:
CONFIRM TARGET DESIGNATION
Duration = 3 sec.
Corrected Duration = __________ sec

	Visual	Auditory	Cognitive	Manual	Speech
% Time*					

TASK:
CONFIRM TARGET IDENTIFICATION/CLASSIFICATION
Duration = 3 sec.
Corrected Duration = __________ sec

	Visual	Auditory	Cognitive	Manual	Speech
% Time*					

* use increments of 25%

Figure 7 -- Resource Load Estimation Form for Part 2 Study

Network Simulation Construction

The network model constructed in the baseline part of this study (described above) was also used in this portion, though only the portion associated with the attack phase was used in this part.

Workload Model

The function used to calculate workload based on the subjective ratings was a simple summation of resources across all tasks being performed at each point in the mission timeline. Total workload at each moment is represented as a vector with five components (corresponding to the five resource channels). Since this workload model defines the threshold for workload only on the basis of individual channels, there is no scheme for aggregating a scalar workload value over the five channels as there was with the W/INDEX formula used in the Part 1 study.

Task Offloading Validation

We attempted to extend the investigation of workload by examining the pilots' judgments of dynamic function allocations for the same mission phase. We developed a software tool to step the subjects through the mission timeline, showing them what tasks were to be performed at each moment. For each timeline time-step, the subject was asked to indicate whether or not he could acceptably accomplish all required tasks without assistance or postponement. In each case where he indicated that he could not perform all tasks simultaneously, he was asked to specify which tasks he would offload, assuming that the offloading would assign the task to an automation capability that was slightly inferior to the pilot's own capability. The software tool which presented the timeline review to the subjects and collected their judgements of task offloading is called the Function Allocation Simulation System (FASS). The principal interface screen for this tool is illustrated in Figure 8.

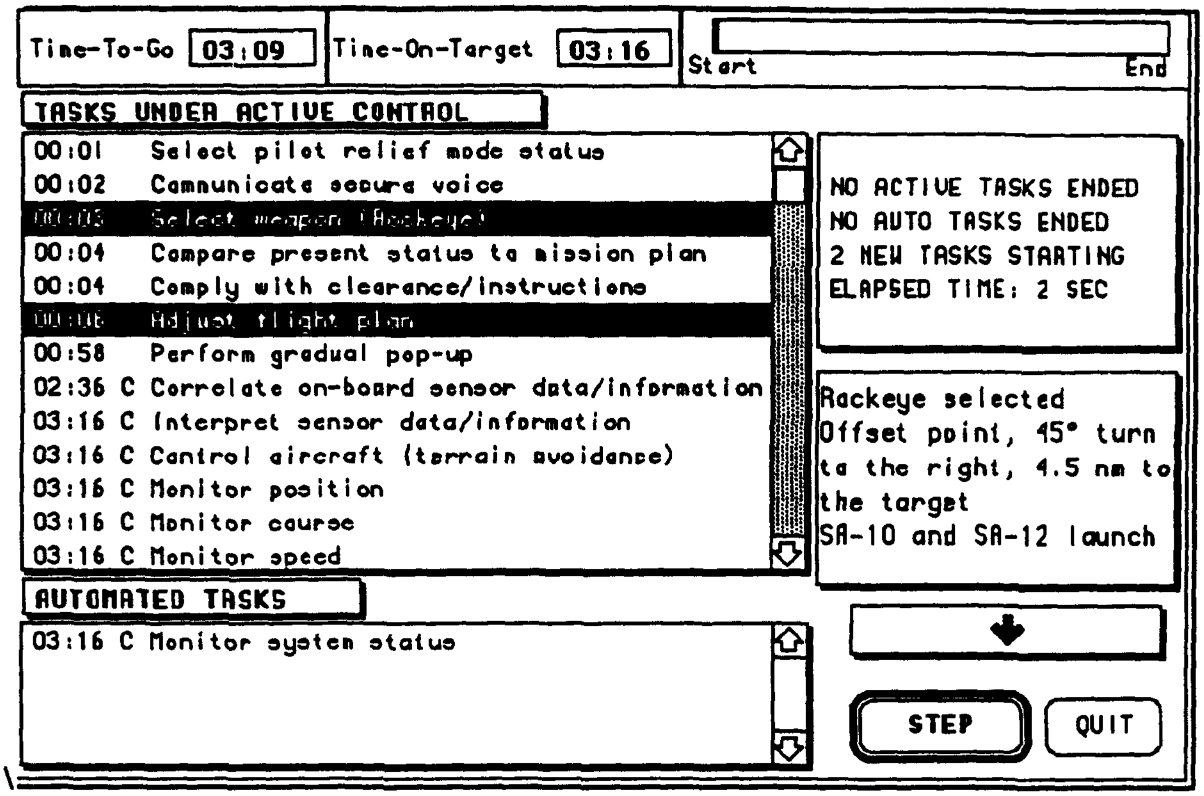


Figure 8 -- Primary Display Screen of the Function Allocation Simulation System (FASS)

PART 2 - RESULTS

Correlations

Means, standard deviations, and correlations of the workload ratings of the seven resources across all tasks were obtained independently for each of the three subjects. These data are presented in Table 2. Although a few significant correlations are evident, there are far fewer high correlations than there were for the case of the seven-channel workload representation used in Part 1 of this study. It is also important to note that these data are based on considerably fewer observations than the analogous data in Part 1 because only the attack phase of the mission was used for this phase of the study. Because of the limited quantity of data in this portion of the study, factor analyses of the data were found to be unstable and unhelpful in the interpretation of results.

Table 2 -- Correlations for Part 2 Study

Resource Channel	Mean	S.D.	Correlations 1	2	3	4	5
Subject 1's Data							
1 Seeing	49.38	23.68		-0.285	-0.541	-0.028	-0.458
2 Hearing	8.13	18.25			-0.213	-0.346	0.505
3 Thinking	52.50	13.63				-0.005	-0.309
4 Manual	25.63	23.68					-0.083
5 Speaking	1.88	8.75					
Subject 2's Data							
1 Seeing	35.00	23.89		-0.322	0.399	-0.056	-0.238
2 Hearing	1.88	8.75			-0.309	0.134	0.892
3 Thinking	38.13	21.17				-0.373	-0.292
4 Manual	18.75	23.85					0.042
5 Speaking	1.25	7.91					
Subject 3's Data							
1 Seeing	20.00	10.13		0.066	0.093	0.030	-0.459
2 Hearing	4.38	9.62			-0.294	-0.382	0.196
3 Thinking	50.00	16.98				0.089	-0.171
4 Manual	19.38	10.57					0.124
5 Speaking	1.25	5.52					

Workload Predictions and Task Offloading Validation

Two of the three subjects seemed to have no problem in making judgements of task offloading as requested, as reflected by the fact that they made many such judgements across the complete duration of the mission timeline. (Subject S2 selected 19 of the 40 tasks for offloading during some portion of the timeline, while subject S3 selected 11 of the 40 tasks in the same fashion.) The third subject (S1), however, had considerable difficulty with this request and chose to perform all tasks manually throughout the entire timeline. In later discussion with this third subject, it became clear that his problem was with the idea of assigning mission-critical tasks to an uncertain, suboptimal automation facility (as postulated in the instructions as the recipient of responsibility for offloaded tasks). It was also clear that this subject fully accepted the use of the many automation capabilities that are currently available in the F/A-18 aircraft and that he would be willing to use

additional automation functions as they were appropriately validated and integrated. Thus, there seems to have been a failure on the part of the experimenters in this case to communicate the focus of this study on task offloading as opposed to assessment of automation options. Accordingly, the remaining analysis in this section will focus only on the results of the two subjects who did seem to make effective judgements of task offloading.

Workload profiles were generated for each subject using a spreadsheet which indicated which tasks were active in each time-step of the timeline. Total resource loading was then calculated, for each subject and each time-step, by simply adding the percentage estimates across all of the active tasks. Two profiles were generated in this fashion for each of the subjects. The first set of profiles indicate the total resource loads that are estimated assuming that no tasks are offloaded, and the second set are based on the pilot performing only the tasks that were not offloaded. These profiles are presented as Figures 9 through 12.

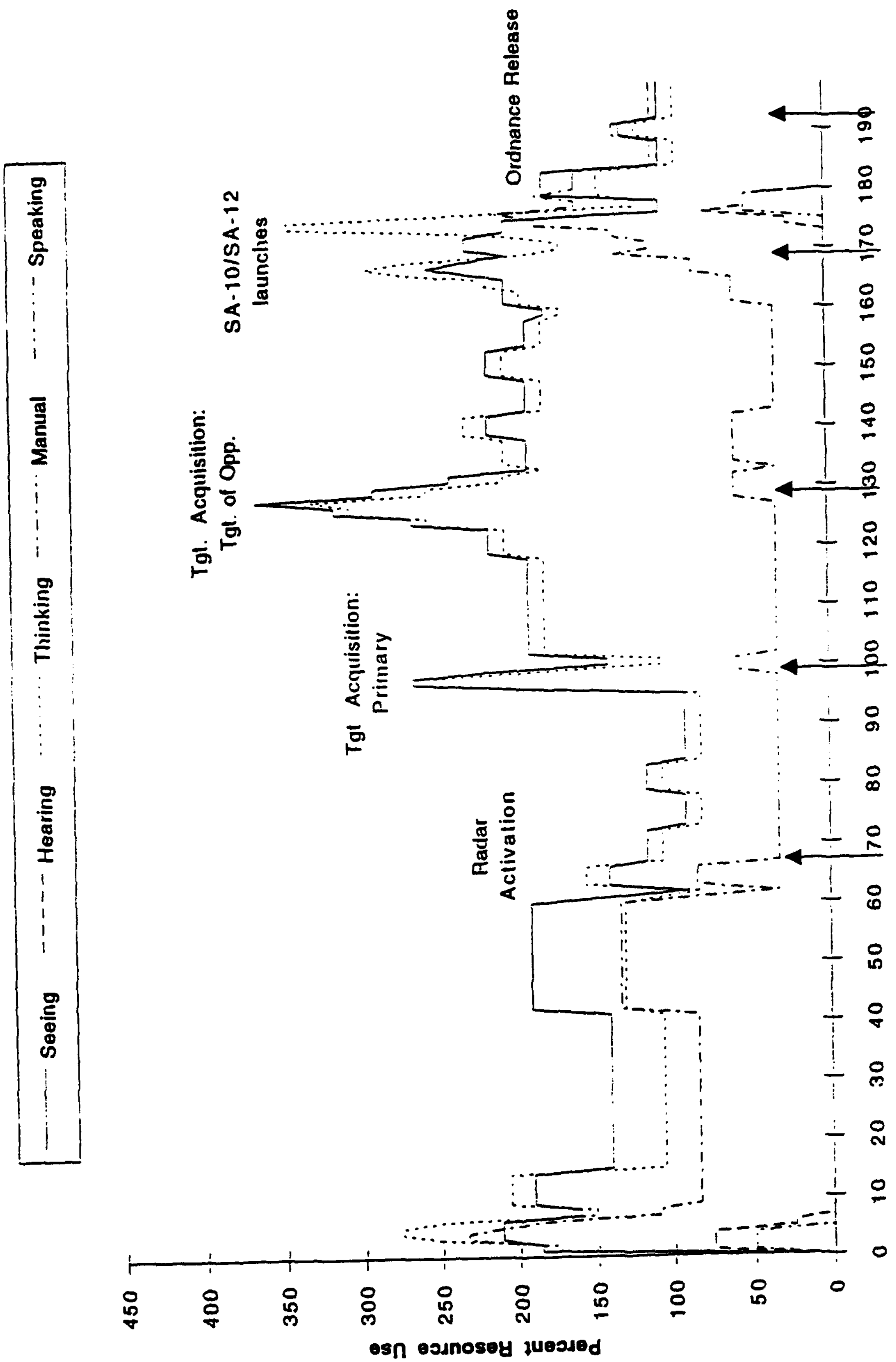


Figure 9. Subject 2 Resource Load Timeline for All Tasks

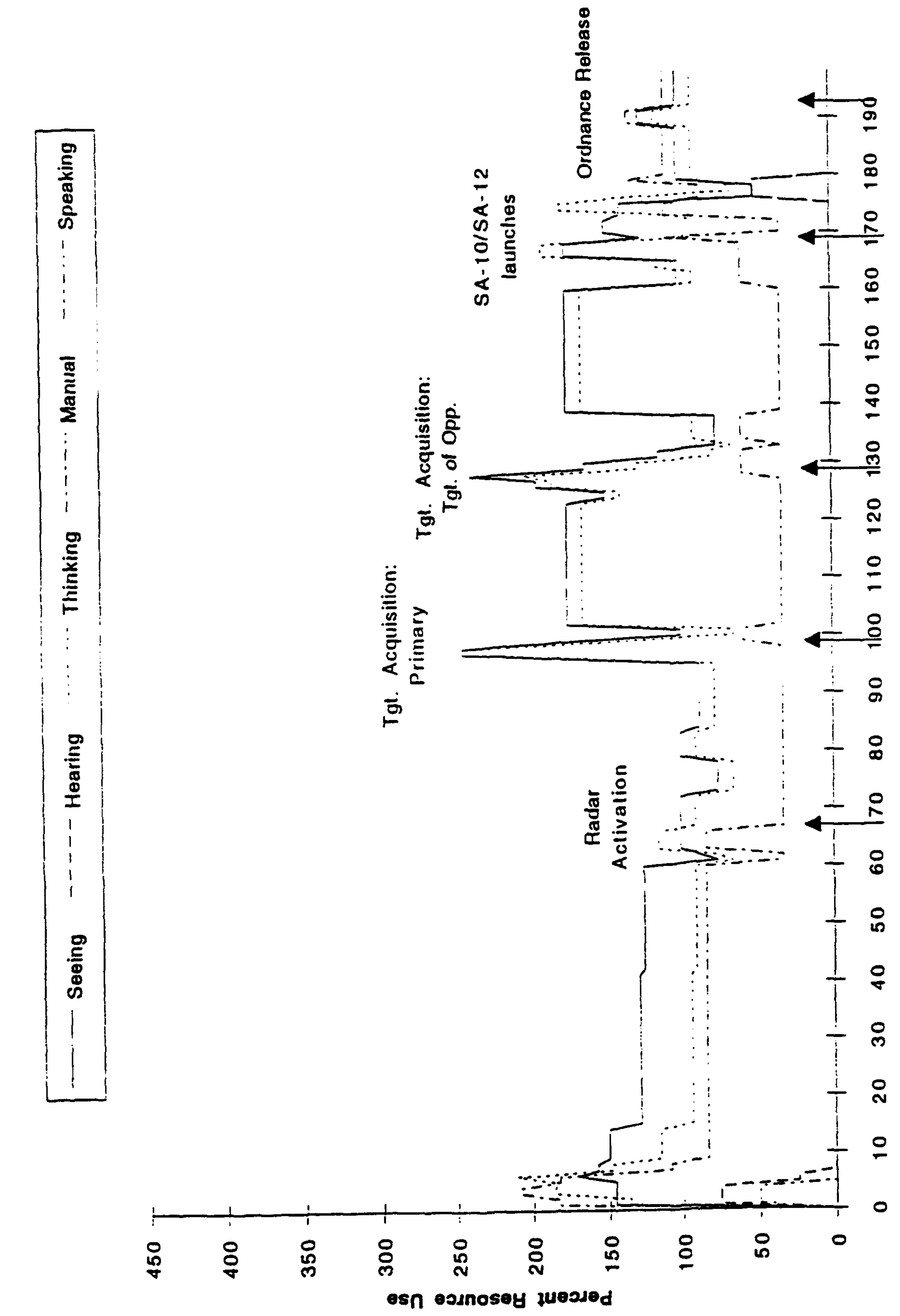


Figure 10. Subject 2 Resource Load Timeline for Retained Tasks

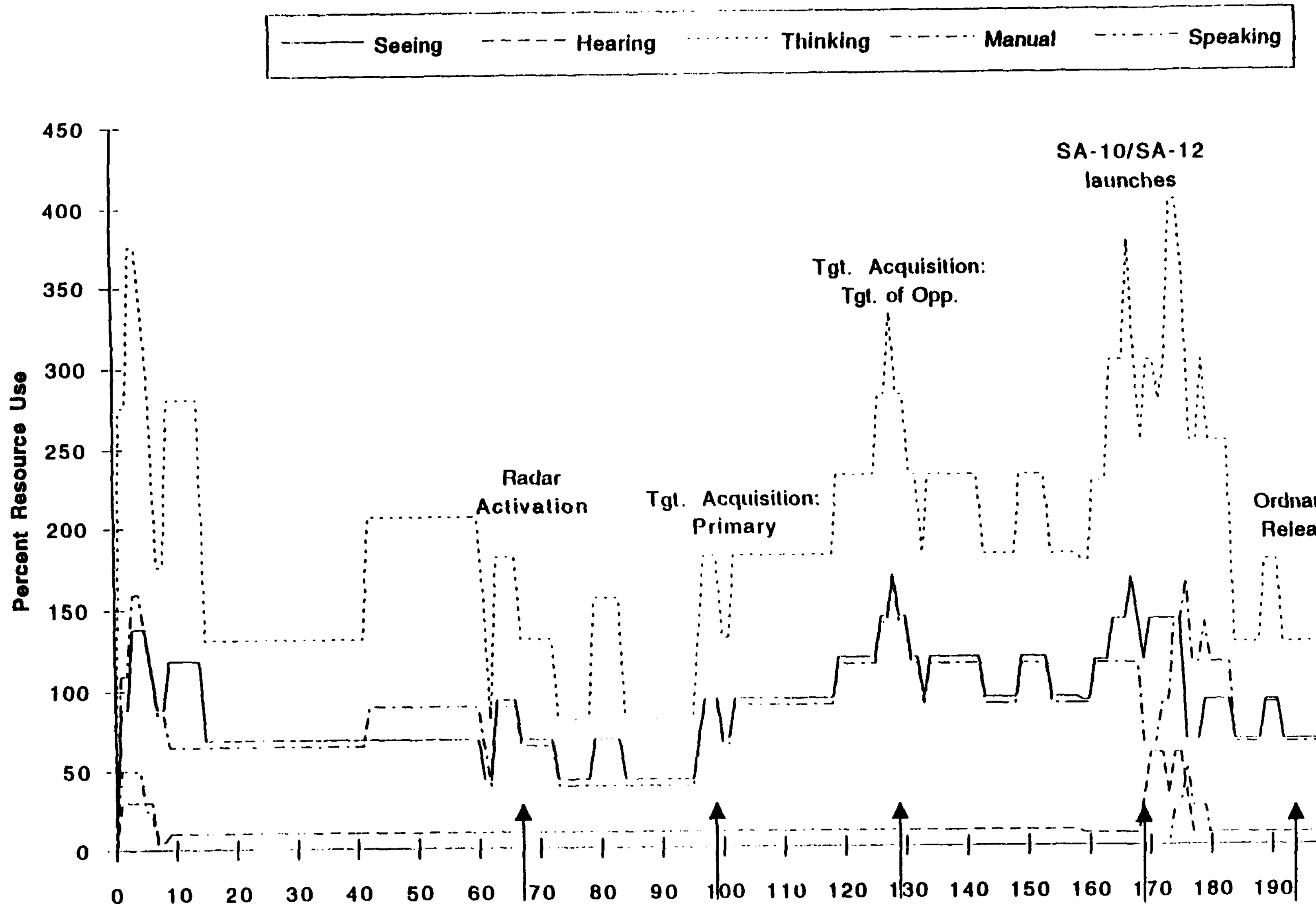


Figure 11. Subject 3 Resource Load Timeline for All Tasks

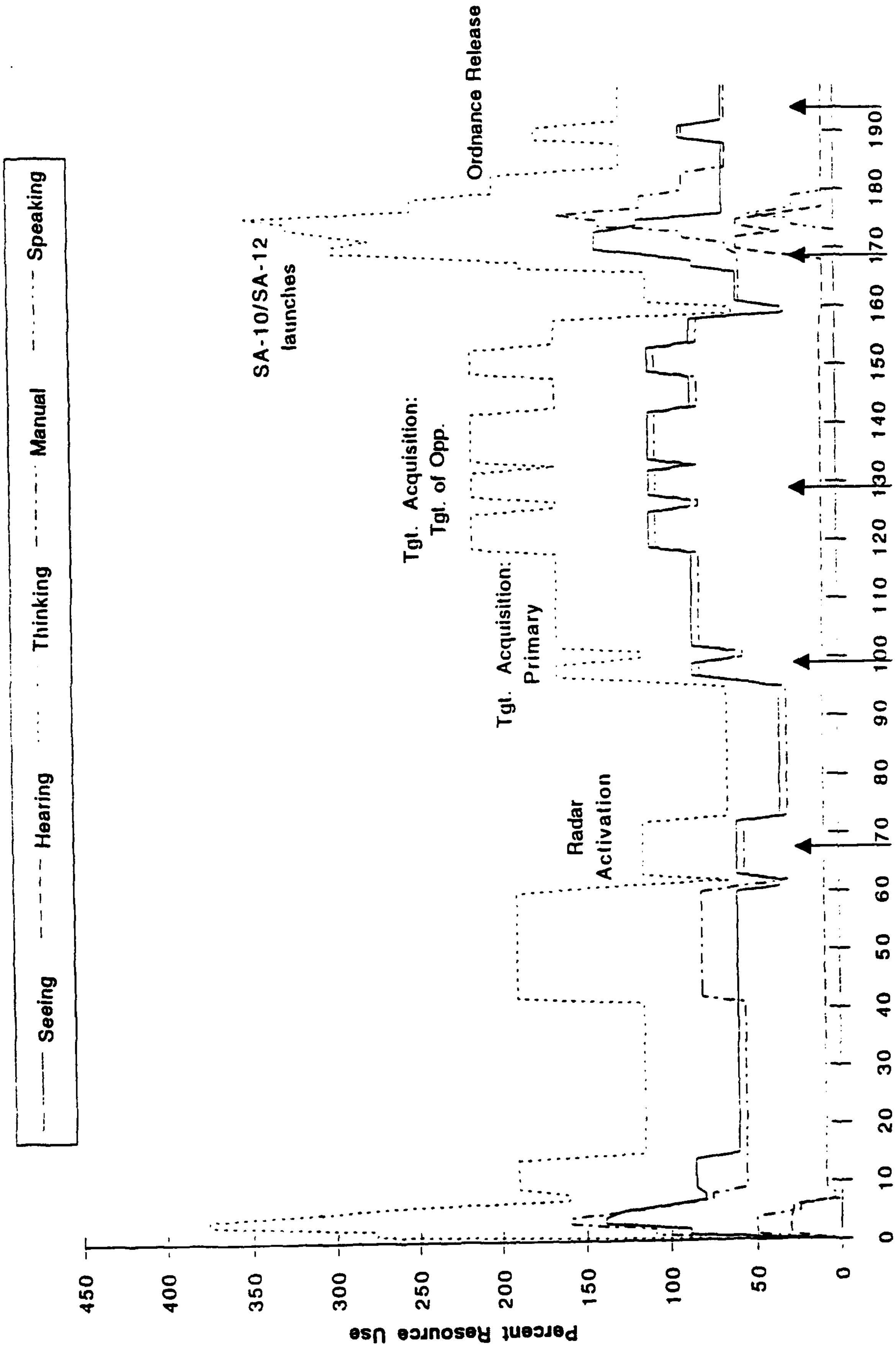


Figure 12. Subject 3 Resource Load for Retained Tasks

DISCUSSION OF PART 2 - TIME-BASED WORKLOAD ESTIMATION

The results of Part 2 of this study suggest some promise for the time-based workload estimation concept. Time-based resource load estimates were readily provided by all subjects for all tasks. Subjects seemed reasonably consistent with one another in the average loads which they assigned to each of the five resource channels. Correlations of load estimates across the channels indicated that the five channels were reasonably distinct from one another. At the same time, it must be recognized that there is a considerable literature documenting problems and biases which people have in estimating time intervals. Indeed, distortions in time estimation abilities have been used specifically to measure workload effects (Hart, 1975; Hicks, Miller, and Gaies, 1977). However, it should also be recognized that these estimation difficulties relate to the estimation of time intervals rather than percentages of resource utilization within a predefined interval. In order to further validate this new technique, it is appropriate to design experimental research to evaluate the abilities of people to make these time percentage estimates. Although it is very difficult to identify time devoted to cognitive activity, it should be possible to identify activities associated with the input and output resource channels.

The attempt to validate the new workload assessment technique using the timeline review method is inconclusive, but encouraging. Since this part of the study focussed exclusively on the mission phase (i.e., Attack) for which the W/INDEX model in Part 1 presented the greatest problem with regard to a threshold, these results for the time-based technique are especially promising. The resource-load timelines for all tasks and for offloaded tasks (Figures 9 to 12) suggest the plausibility of 100% as the limiting threshold. The two subjects who made offloading decisions effectively moved both the maximum and average levels of resource loads for the heavily loaded channels (i.e., seeing and thinking) closer to 100%, though the levels for these channels were still between 100% and 200% for much of the timeline (see Figures 13 to 16). Brief excursions above 100% do not necessarily pose much of a problem, as they could potentially be removed by readjusting the periods of resource utilization within the tasks (i.e., by relaxing the homogeneity assumption). Longer durations of resource load above 100% suggest the need for some revisions to the technique, possibly recalibrating individual resource loading scales (e.g., on the assumption that each individual has a different standard for the upper limit of resource capacity, which is not

necessarily 100%) or possibly improving on the timeline review procedure. It should also be noted that all subjects had some objection to working with a pre-established timeline of tasks, and each subject disagreed with the estimated task durations for some tasks. These disagreements certainly created some problems in the generation of the resource time percentage estimates, since the subjects were instructed to maintain the pre-established task times.

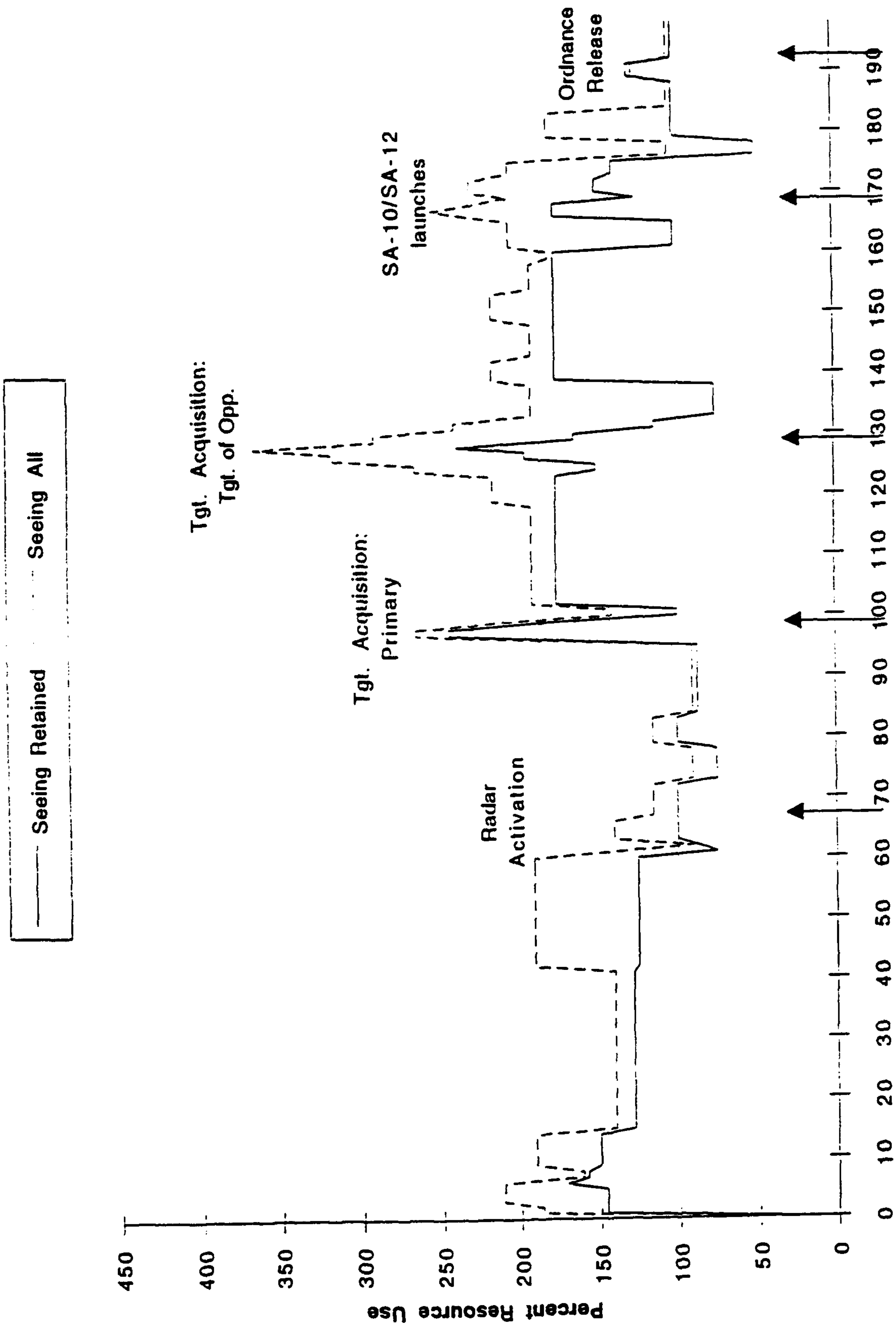


Figure 13. Subject 2 Seeing Resource Load Timeline (All vs. Retained)

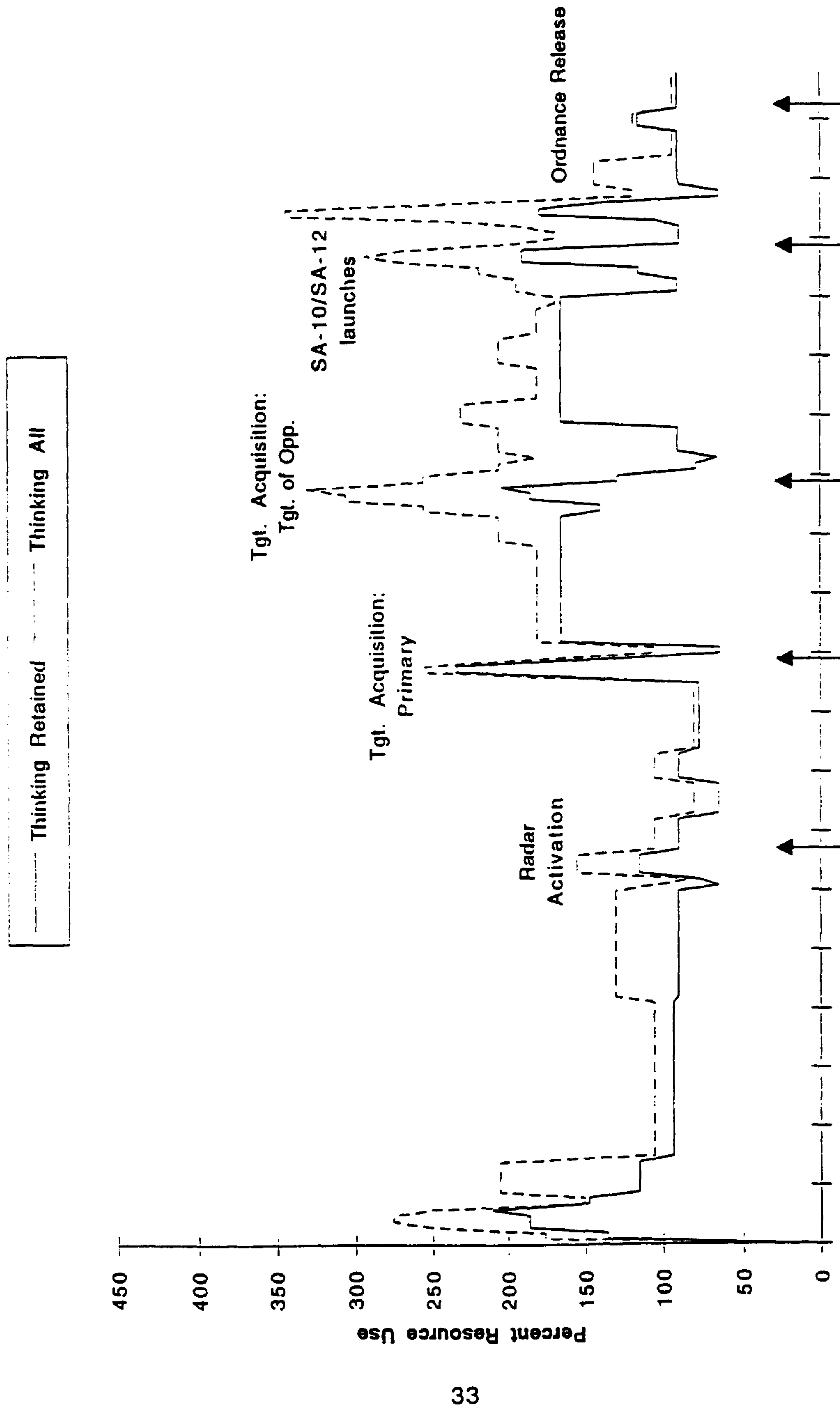


Figure 14. Subject 2 Thinking Resource Load Timeline (All vs. Retained)

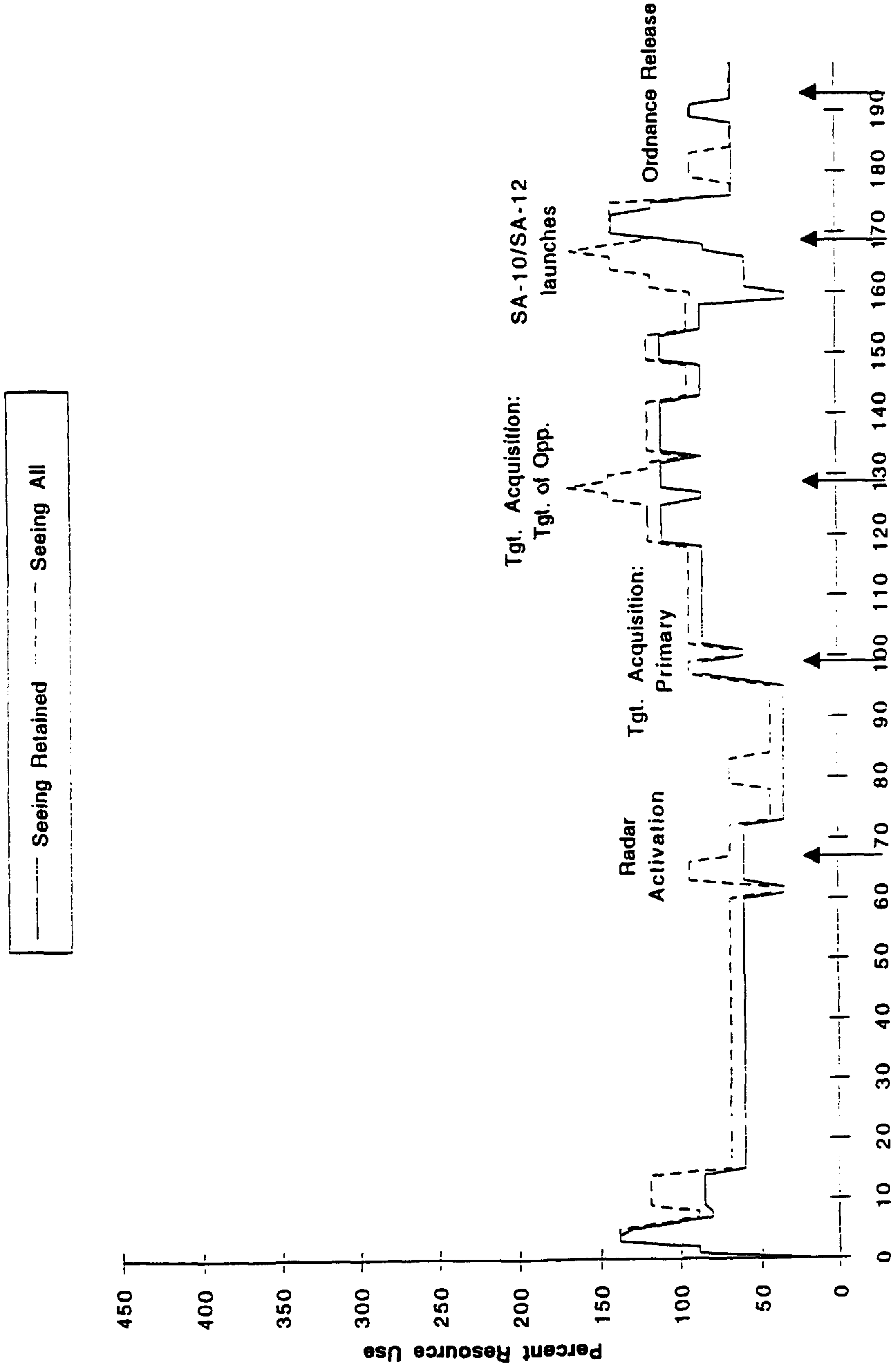


Figure 15. Subject 3 Seeing Resource Load Timeline (All vs. Retained)

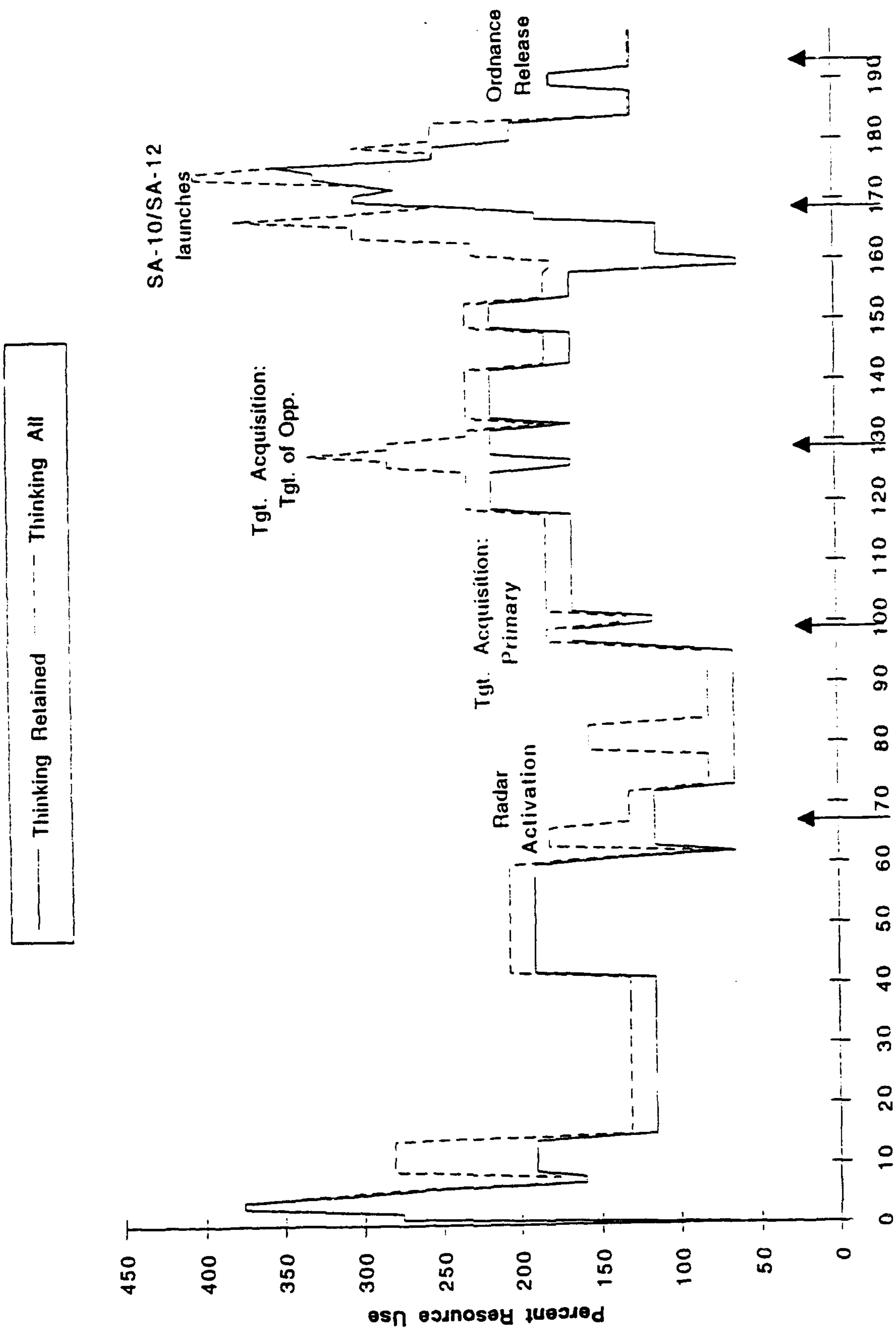


Figure 16. Subjec. 3 Thinking Resource Load Timeline (All vs. Retained)

CONCLUSION

The complexity of the W/INDEX formula (its workload model) and its utilization of conflict matrices certainly give it the appearance of a carefully constructed and precise instrument for determining workload. When the W/INDEX model is further coupled with a task network simulation program, together they can produce a variety of apparently sophisticated outputs (e.g., total instantaneous workload, individual channel loadings, etc.) which, while costly to achieve, may not provide the diagnostic utility they purport to yield. Before these types of prospective workload estimation techniques become widely adopted, we need studies demonstrating that early projected estimates of efforts required for system tasks do, in fact, correlate highly with actual efforts required by those same tasks. This study did not attempt to do this since the system we studied has yet to be developed.

In adopting multiple resource theory as part of the W/INDEX model, the workload rater is asked to go beyond describing overall effort required by a task and, instead, describe the effort levels required for a variety of different perceptual, cognitive, and response activities. Our data suggest that raters, when evaluating systems that have yet to be developed, are limited in their abilities to distinguish separate performance resources that might be required, especially in the cognitive domain. Further studies would also be useful to determine the extent of correlation between both projected times and effort levels and (once the system is developed) the actual times and subjective effort levels expended for each of the resource channels.

We recognize that the concept of workload is broader than the concept of performance time and accuracy. With workload we desire to know how close we are coming to overloading the capacity of the operator rather than simply if the operator will be able to perform all of the assigned tasks. If multiple-resource approaches are to be taken with regard to estimating overall task effort and in discriminating among different types of activities which lead to operator overload, then it seems reasonable to first enquire as to the percentages of overall allocated task times that must be dedicated to each activity type. We offered a time-based workload assessment concept in Part 2 of this study, along with a limited demonstration of a technique for its validation.

Much still remains to be done to fully develop and evaluate these new concepts and techniques. The dependence of the workload estimations on a pre-

established task timeline poses a significant problem since it requires someone to estimate task durations and we know that people have difficulty in making such estimates. However, this problem is common to all projective, task-network-based assessment techniques. Studies to evaluate peoples abilities to make reliable estimates of resource utilization are warranted, for example correlating eye movement or hand movement records with estimates of visual or manual resource loads. Refinements to our validation technique based on timeline review are also needed, including improvements in communicating to the subject the specific character of the tasks to be performed at each moment. One candidate approach for achieving greater fidelity in this kind of validation process is to have the subjects perform the tasks in some type of flight simulator with the option of selecting tasks for offloading at any time by pressing easily accessible buttons. It would then be possible to compare both the resource load estimates and the analytic task-offloading judgements with the actual real-time decisions made in the course of the simulated mission. This type of simulator-based validation of the new workload assessment technique is coincidentally also the baseline concept for adaptive automation in the cockpit (designated as "pilot initiative" invocation of automation), which is currently being studied through two related programs at the Naval Air Warfare Center, Aircraft Division, Warminster.

Although this research began with a focus on providing support for conventional function allocation decisions in cockpit design, it has become increasingly evident that some of its greatest benefits may lie in its application to the domain of adaptive automation. Accordingly, we will conclude this section with an overview of the ongoing work at the Naval Air Warfare Center in the area of adaptive automation.

In the past, most automation designs, in aviation as well as other industries, were technologically driven. Engineers automated whatever they were able to automate on the basis of available gadgetry, simply assuming overall system performance would improve (Morrison, Gluckman and Deaton, 1991). This lack of concern for the human operator as an effective element within this system, however, led to a variety of automation-induced errors and concerns (Chambers and Nagel, 1985; Parasuraman, 1987; Parasuraman, Bahri, and Molloy, 1991; Wiener, 1977; Wiener and Curry, 1980; Wiener, 1988). As a result, interfacing the human operator with his/her automated cockpit became a major impetus for many aviation human factors specialists.

Wickens and Kramer (1985) discuss three major types of automation that may be implemented in human-computer systems: "automation that assists", "automation that replaces", and "adaptive automation" (p. 335). The first two types (automation that assists and automation that replaces) are more traditional or "static" forms of automation, in which specific functions are allocated to human and automated components early in the design process, and the consequent roles remain relatively unaltered by varying situational concerns. Furthermore, invocation of the automation (turning it on or off) is a responsibility of the human operator. Adaptive automation, on the other hand, is implemented in a dynamic manner, so that the functions allocated to the human and automated components change with the changing demands and characteristics of the system. Furthermore, the method of invocation is viewed as an additional concern in the "sharing" of responsibilities between the human and machine information processing elements of the system. As described by Morrison, Gluckman and Deaton (1991), "adaptive automation is automation which is capable of engaging and disengaging itself in response to either 1) the occurrence of a critical event or events, or 2) based on the performance of the human component(s) in a person-machine system" (p. 1).

It is the primary purpose of the Adaptive Automation for Intelligent Cockpits (AFAIC) and Adaptive Invocation Development (AID) programs at the Naval Air Warfare Center, Aircraft Division, Warminster to determine the benefits of such strategies, and to modify , invent, and recreate appropriate strategies where possible.

Wickens (1984) suggests three potential benefits resulting from automation in general. These include the allocation to automation of those functions that are potentially dangerous to humans and/or those which humans cannot do; those activities humans often perform poorly due to overloading or underloading of processing capacity; and, finally, those tasks needed "to supplement or augment human perception, memory, attention, or motor skill" (p. 334).

These potential benefits of automation cannot be assumed as they were in the past. They must be evaluated from the perspective of the human-machine system. Before designers can implement the optimal automation strategy (or strategies) for a particular environment and situation, they must understand the information-processing system for which the benefits are intended. A vital component of this information processing environment is the human operator (i.e. pilot). In order for this to be accomplished, reliable and informative evaluative

techniques concerning human performance and information processing issues are critical.

Although "workload" is both an all-encompassing and a somewhat evasive term, it is relatively useful in communicating concepts within the cognitive and human factors disciplines. The measurement of workload, however, is far more difficult to grasp than its face-value comprehension. One program addressing a variety of the issues, both positive and negative, in the assessment of workload has been discussed in this paper. Such programs, which concentrate all efforts upon maximizing both the quality and quantity of information which can be attained from human assessments, are important to exploratory developmental programs, such as AFAIC and AID. In these latter programs, human factors designs are based upon the development and testing of theoretical concepts concerning issues, such as workload and situational awareness, which are thought to be related to human performance within adaptively automated aviation.

Two philosophies regarding changing automation status are currently being studied: Critical Event Centered and Human Performance Centered. In the former case, external events, an example of which could be increased task loads, affect the decision to adaptively automate. The Critical Event Centered philosophy is one in which the potential exists to design the algorithm determining automation early in the process. Such an algorithm might be based, for instance, on research indicating that under certain task load situations, the human operator becomes overloaded, and his/her performance drops if certain tasks are not adaptively automated. It then becomes obvious why programs concentrating upon the development of accurate projective workload (and related) assessments are crucial to adaptive automation research.

Two other aspects of the AFAIC Taxonomy, Strategy and Decision Stability, illustrate the importance of the particular approach taken in the projective assessment program discussed in this paper. This approach is one in which attempts at projective workload assessments examine several issues within the domain of time and resource load techniques. Preliminary research within the AFAIC program suggests that designers should be closely examining the unique cognitive nature of required tasks before choosing the adaptive automation strategy that would produce optimal system performance. In particular, a task dichotomy has been hypothesized on the basis of decision stability, discriminating tasks associated with a stable versus a dynamic internal model. In the case of stable internal models, the information required to make accurate task decisions,

once learned, does not change across time (Carmody and Gluckman, 1993). In the case of unstable internal models, on the other hand, such information does change across time. Research within the AFAIC program has indicated that these aspects of the task dichotomy, coupled with various automation strategies, differentially affect subject workload and situational awareness, thereby differentially affecting performance. Preliminary results from research on adaptive automation and workload, in particular, suggests that coupling various automation strategies with tasks of characteristically different decision stability has a measurable effect on workload. However, this effect is complex, and could greatly benefit from a program dedicated to determining the critical resource elements involved in workload, as well as the best method for measuring overload. It is a hope of the AFAIC program that such workload effects, properly measured and in conjunction with other performance issues, could be used early in the design process to determine optimal automation strategies on the basis of task type.

REFERENCES

Bierbaum, C.R., Fulford, L.A., & Hamilton, D.B., (1989). *Task analysis/workload (TAWL) user's guide.* (Technical Report ASI-690-323-89(a)). Fr. Rucker, AL: Anacapa Sciences, Inc

Card, S., Moran, T., & Newell, A. (1983) *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carmody, M.A. and Gluckman, J.P. (1993). Task specific effects of automation and automation failure on performance, workload, and situational awareness. In R.S. Jensen (Ed), *Proceedings of the Seventh International Symposium on Aviation Psychology. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory (in press).*

Chambers, A.B. and Nagel, D.C. (1985). Pilots of the future: Human or computer? *Communications of the Association for Computing Machinery,* 28, 1187-1199.

Cohen, D. (1990) *Identification of advanced technology crew station decision points and information requirements* (Report No. NADC-90XXX-60) Warminster, PA: Naval Air Development Center.

Glenn, F., Cohen, D., Barba, C., & Santarelli, T. (1990). *The advanced technology crew station: Initial workload assessment* (NADC Technical Report No. NADC-90117-60). Warminster, PA: Naval Air Development Center.

Hart, S.D. (1975, May). Time estimation as a secondary task to measure workload. *Proceedings, 11th Annual Conference on Manual Control* (NASA TMX-62, N75-33679, 53), pp. 64-77. Washington, DC: U.S. Government Printing Office.

Hart, S.G. & Staveland, L.E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.S. Hancock and N. Meshkati (Eds.), *Human mental workload.* Amsterdam: North-Holland, pp 139-183.

Hicks, R.E., Miller, G.W., & Gaies, G. (1977). Concurrent processing demands and the experience of time in passing. *American Journal of Psychology, 90*, 431-446.

Lane, N., Strieb, M., & Wherry, R. (1977) "The human operator simulator: workload estimation using a simulated secondary task." In *Methods to Assess Workload*. NATO/AGARD Conference Proceedings CP-216.

Lane, N., Strieb, M., Glenn, F., & Wherry, R. (1981) "The Human Operator Simulator: An Overview" *Manned Systems Design: Methods, Equipment, and Applications*. J. Moraal and K.-F. Kraiss (eds.). New York: Plenum Press.

Laughery, K., Dahl, S., Kaplan, J., Archer, R., & Fontenelle, G. (1988) A manpower determination aid based upon system performance requirements. In *Proceedings, 32nd Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society.

Laughery, R., Drews, C., & Archer, R. (1986) A MicroSAINT simulation analyzing operator workload in the LHX helicopter. In *Proceedings of the NAECON 86 Meeting*, Dayton, OH.

McCracken, J.H. & Aldrich, T.B. (1984, June). *Analysis of selected LHX mission functions: Implications for operator workload and system automation goals* (Technical Note ASI479-024-84). Fort Rucker, AL: Anacapa Sciences, Inc.

Morrison, J.G., Gluckman, J.P. and Deaton, J.E. (1991). *Program Plan for the Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program*, (Final Report No. NADC-91028-60). Warminster, PA: Naval Air Development Center.

North, R. & Riley, V. (1989). W/INDEX: A predictive model of operator workload. In G. MacMillan, D. Beevis, E. Salas, M. Strub, R. Sutton, & L. Van Breda (Eds.), *Applications of human performance models to system design*. (pp. 81-89). New York, NY: Plenum Press.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29(6), 695-706.

Parasuraman, R., Bahri, T., and Molloy, R. (1991). *Adaptive automation and human performance: I. Multi-task performance characteristics.* (Technical Report No. CSL-N91-1), Cognitive Science Laboratory, The Catholic University of America, Washington, D.C.

Reid, G., Shingledecker, C., & Eggemeier, T. (1981) Application of conjoint measurement to workload scale development. In *Proceedings, 25th Annual Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Veda, Inc. (1990) *ATCS mission timelines.* (Veda Report No. 33236-90U/P3838). Warminster, PA.

Weiner, E.L. (1977). Controlled flight into terrain accidents: System-induced error *Human Factors*, 19(2), 171-181.

Weiner, E.L. (1988). Cockpit automation. In Weiner, E.L. and Nagel, D.C (Eds), *Human Factors In Aviation.* San Diego, CA: Academic Press, Inc.

Weiner, E.L. and Curry, R.E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.

Wickens, C. (1984). *Engineering Psychology and Human Performance.* Columbus, OH: Merrill.

Wickens, C.D. and Kramer, A. (1985). Engineering psychology. *Annual Review of Psychology*, 36, 307-348.

APPENDIX A FASS SOFTWARE DESCRIPTION

FASS provides a low fidelity simulation of a man-in-the-loop system and collects, from domain experts, *subjective estimates of cognitive workload*. The initial development of FASS utilized a timeline of tasks representative of the tasks performed by U.S. Navy pilots in the 'attack' phase of a 'strike' mission. This one phase of a longer mission was selected because of the relatively high number of concurrent tasks being performed and the expectation that high workload values would be present. As described earlier, estimates were collected across five channels -- seeing, hearing, thinking, manual, and speaking. The following sections document the source data (task timeline), output data (task automation requests and sensory channel loadings), and the design of the FASS software.

Main Simulation Screen

Figure 8, presented earlier, shows the primary FASS window which includes a variety of fields, buttons, and graphic elements. Fields and graphics provide feedback to subjects on the status of the simulation. Subjects use buttons to initiate *automation decisions* and control their progress through the simulation. Feedback is provided in the fields labeled 'Time-To-Go', 'Time on Target', 'Tasks Under Active Control', and 'Automated Tasks'. Two additional fields are displayed to the right of the list of tasks under active control which indicate, respectively, changes at each time step to the task lists (such as how many tasks have started and ended) and milestones in the mission (such as weapon selections, weapon launches or flight maneuvers). In the upper right corner of the main FASS display a slider provides a graphic representation of progress through the mission.

Subjects in the experiment interact with the simulation through the buttons labeled 'Step' and 'Quit' and an up/down arrow button, or by selecting a line in the fields displaying task data. When a task is selected in the 'Tasks Under Active Control' field, the arrow button becomes a down arrow (↓) and clicking on the button moves the task to the list of automated tasks. Conversely, when a line is selected in the Automated Tasks field, the arrow button becomes an up arrow (↑) and clicking on the arrow button moves the selected task to the 'Tasks Under Active Control' field. When a task is automated, the system displays a screen to collect further justifications from the subject regarding the workload in that task. This

screen, as well as the data collected, is described later in the section 'Automation Rationale'.

The only other screen in FASS is a logon window which collects the name of the subject and allows selection of a mission and phase. This initial experiment implemented only the 'attack' phase of a 'strike' mission but other timelines could easily be adapted and incorporated for use in FASS.

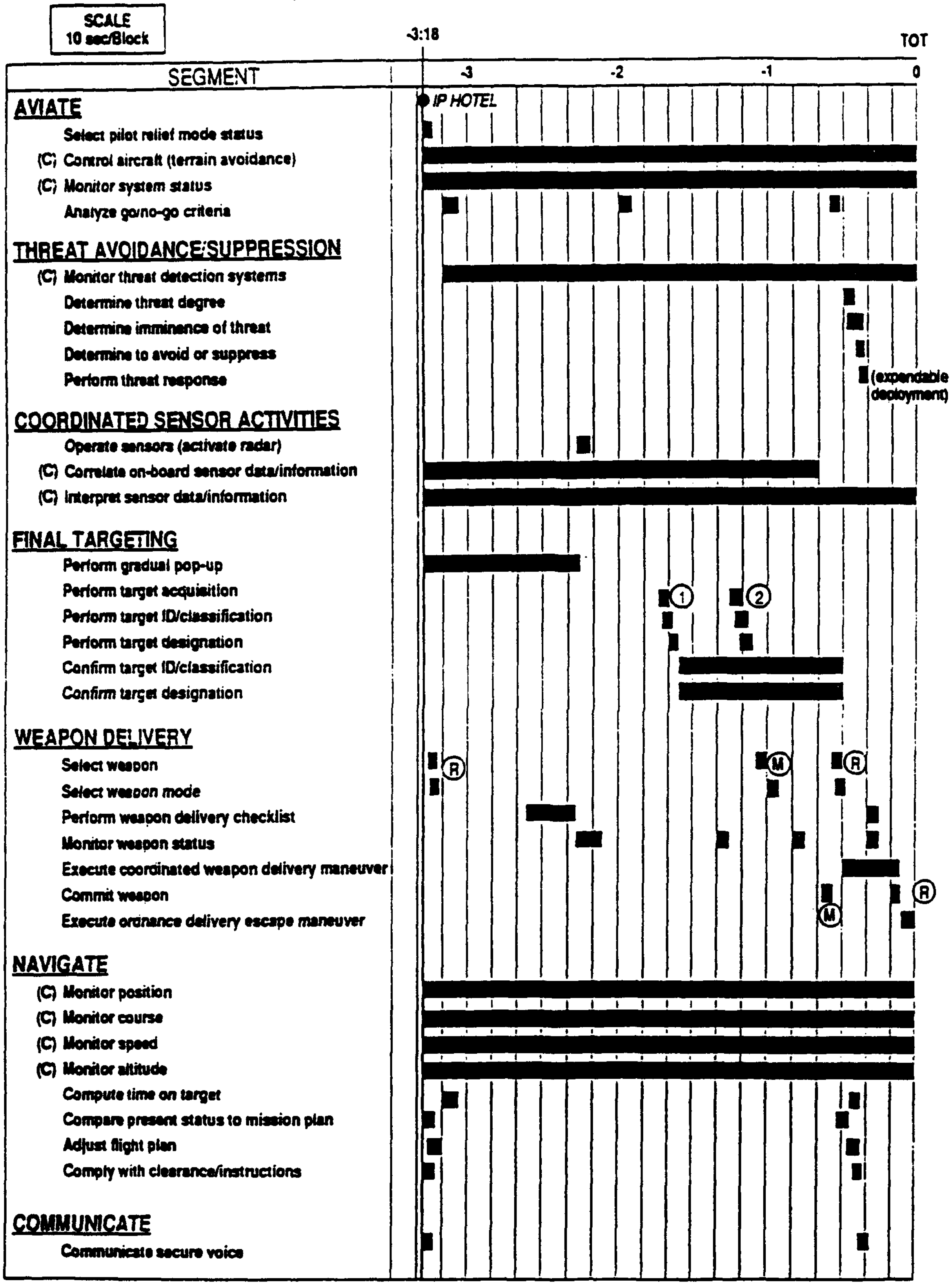
Input and Output Data

An important aspect of control in the experimental design is the timeline of tasks. The timeline is static and constant in that the same tasks are presented in the same order with the same duration at the same time step each time the simulation is executed. If the timeline were variable, there would be less comparability between the automation judgements made by different subjects. In order to be used in the software, the chart shown in Figure A-1 was transformed into a series of data lines including the duration for the task in seconds, a representation of the initial duration for the task formatted as 'mm:ss', and the task name. All the tasks were stored in an array with the array index corresponding to the start time of the task. That is, tasks starting at time step 1 were stored in array index 1. Tasks initially displayed at the beginning of the simulation, timestep 0, were stored separately and are displayed by an initialization routine. Table A-1 shows a few entries in a timeline data array. Notice that it is possible for multiple tasks to start at the same point in the timeline. A similar method was used to store, locate, and display Mission Milestones and the 'mm:ss' formatted data displayed in the 'Time-To-Go', and 'Time-On-Target' fields.

Snapshot of Allocations

The main data collected from subjects executing the simulation are the task allocations. When a subject clicks on the 'Step' button, the current allocation of tasks in the automated and active control fields are collected and appended to an external data file. The simulation then proceeds to the next time where there is a task start, task end, or mission milestone. During development of FASS, it was decided that subjects should not be making second-by-second allocations of tasks, but rather managing the mix of automated and manually controlled tasks only when there was a change in the tasks currently displayed (i.e., a new task starts or a task ends).

DEEP AIR SUPPORT MISSION: ATTACK PHASE
(3 minutes, 18 seconds)



KEY
C - Continuous
(1) vs. APC column
(2) vs. MBT
(R) Rockeye
(M) I²R Maverick

Figure A-1 -- Deep Air Support Mission: Attack Phase Timeline

Table A-1 Transformed Timeline Data

Array Index /Timestep	Task Starts at This Timestep
1	3 0 00:03 Select weapon (Rockeye), 6 0 00:06 Adjust flight plan
2	3 0 00:03 Select weapon mode (Rockeye)
3	
4	
5	
6	
7	6 0 00:06 Analyze go/no-go criteria, 6 0 00:06 Compute (TOT) time-on-target, 190 0 03:10 C Monitor threat detection systems

char 1 to first space --> duration in seconds
char after first space to second space --> flag for cyclic task (not used)
char after second space to comma or end of line --> string for display which contains 2 or 3 terms which include: initial duration formatted as mm:ss, character 'C' as indicator of continuous task when necessary (or blank, if not), and task name.

Automation Rationale

To execute an automation decision, the subject selects a task from the 'Tasks Under Active Control' list and clicked on the down arrow button. The system then displays the dialog box shown in Figure A-2. This screen presents a check box for the five resource channels and requests that the subject flag those channels for which workload will be reduced by the automation assignment. A second file created during a session with the FASS software records the judgements of which resource channels contributed to the lessening of workload with each assignment to automation.

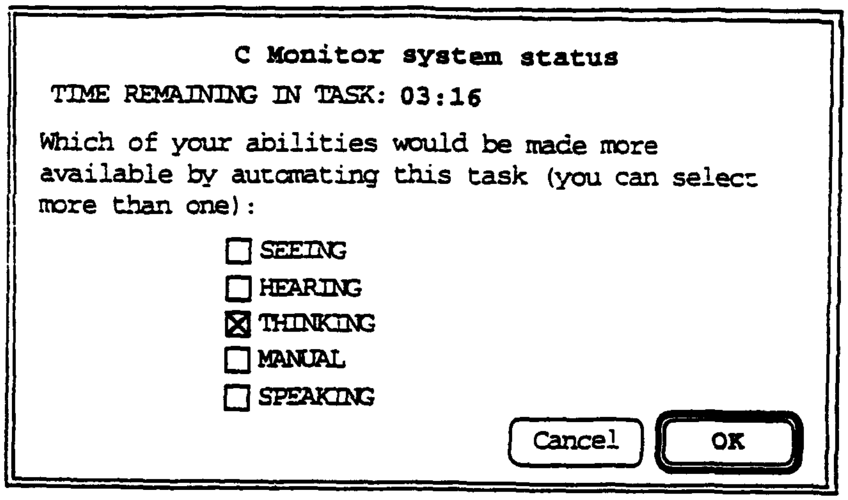


Figure A-2 -- Resource Loadings Dialog Box

FASS Software Architecture

This section further describes the FASS program architecture and purpose of the main code modules. The architecture of the FASS software is highly distributed and non-sequential, mirroring the organization of Supercard. In Supercard (and FASS) code modules called scripts are attached directly to interface components (i.e. buttons, fields, and graphics). Supercard can be considered an object-oriented environment because of its use of objects to build and define graphical user interfaces and their functionality. However, this leads to distributed code without the necessity of defining a main loop or procedure from which all other procedures are called. Because of the non-sequential nature of the software this section discusses in general terms the response of interface components to user actions. Any code listed here would have to be interpreted as a small part of a larger whole to gain a full understanding of the application.

Primary Code Modules

The state transition network in Figure A-3 represents the screens described above as rectangles with arrows between the boxes representing actions which cause other screens to be displayed. The main code modules of the FASS system are located in button scripts represented by the lines connecting rectangles on the STN. The first screen (the 'FASS Startup Screen') includes a path to the 'Enter Subject Name' screen and a 'Cancel' button to exit the simulation.

The 'Enter Subject Name' screen accepts any combination of uppercase letters, lowercase letters, and numbers to specify the identity of a subject in the experiment. Buttons labeled 'OK' and 'Cancel' provide a path to continue the experimental trial or to return to the previous screen. The 'OK' button executes code that displays the 'Main FASS Display Screen' and the set of tasks initially active at the start of the simulation.

On the 'Main FASS Display Screen' there is code distributed among several interface objects, buttons, and fields, which responds to user actions. Specifically, code in the 'Active Tasks' field and the 'Tasks Under Active Control Field' perform complementary actions such as selecting a line in the target field, deselecting text in the other field, and changing the direction of the arrow button.

The code encapsulated in the arrow button handles movement of a task between the fields and is sensitive to the direction of the arrow. When the arrow

button moves a task to the automated tasks area, the scripting displays the automation rationale screen to capture further information from the subject.

The 'Step' button on the 'Main FASS Display Screen' initiates a series of actions which moves the simulation to the next decision point in the simulation. This involves decrementing the remaining time displayed on each task line, and updating the simulation status fields 'Time-To-Go' and Time-On-Target', displaying the summary of changes and, if necessary, displaying any milestone. Code encapsulated in the 'Step' button is also responsible for determining that the end of the simulation has been reached so that external data files can be closed and the simulation exited.

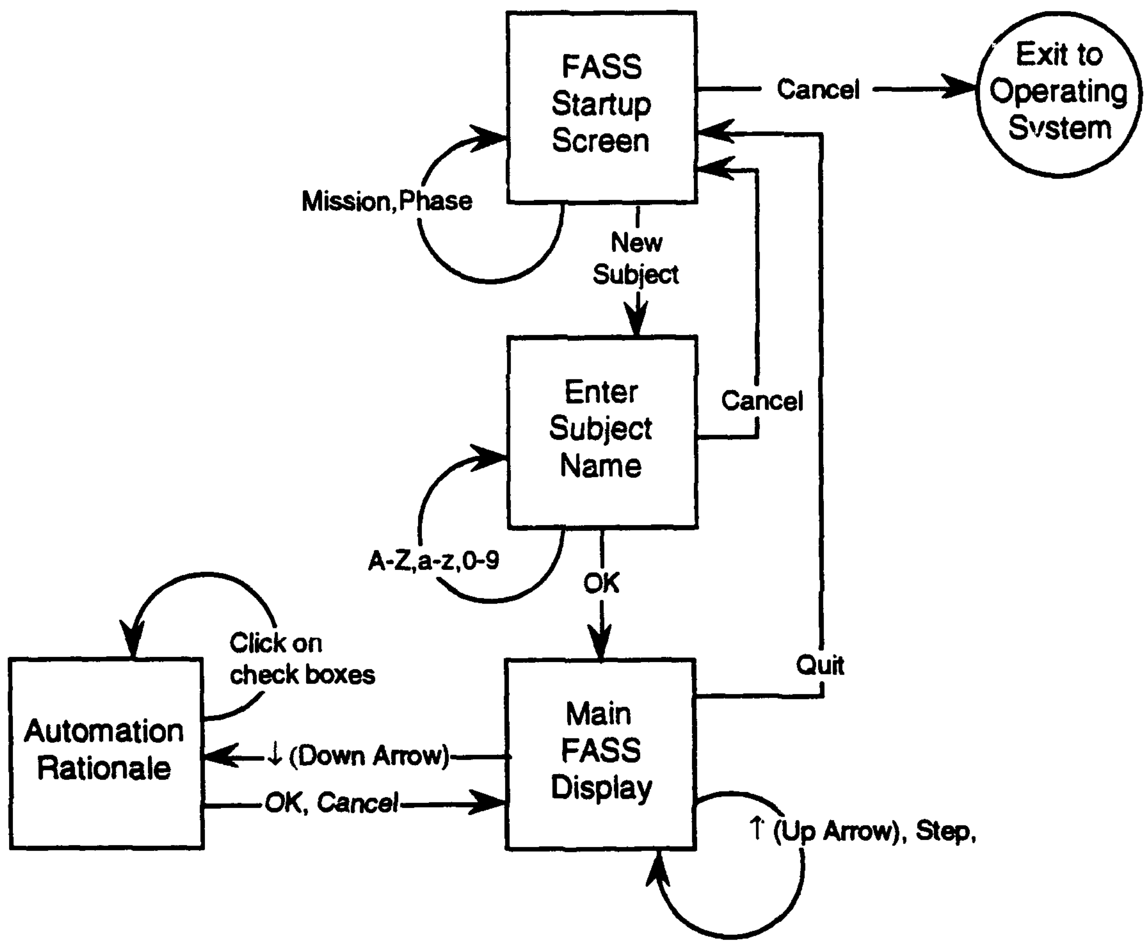


Figure A-3 FASS State Transition Network

Experiences

Our experience in implementing FASS yielded several conclusions regarding the suitability of Supercard for development of prototype systems and for use in operationalizing an experimental design. Supercard is capable of the full scale development of mouse-driven, direct manipulation software, and the Supercard programming language, Supertalk, provides a reasonable set of functions for controlling interface objects, storing and manipulating data, etc. However, because Supercard uses an interpreted programming language, the speed of performance of FASS is barely adequate. FASS does not attempt to present the tasks in real time, nor are pilot subjects asked to actually perform tasks as could be the case in a more realistic simulation. Supertalk also uses weak typing of variables which can be an advantage in some cases, but puts a burden on the programmer or implementer to track the use and expected contents of variables. Initial design and continued development of FASS was accomplished quickly and cheaply because the compile-test-debug cycle is shorter in an interpreted programming environment. However, the software still required significant testing to be used reliably in an experimental design.

DISTRIBUTION LIST

	No. of Copies
Dr. Jim Ballas Naval Research Lab Code 5534 4555 Overlook Avenue, SW Washington, DC 20375-5000	1
LCDR John Deaton Defense Training and Performance Data Center 3280 Progress Dr. Orlando, FL 32826-3229	10
Dr. Peter Hancock 172 Pillsbury Dr. SE 164 Norris Hall Minneapolis, MN 55455	1
CDR S.D. Harris Naval Air Systems Command Code AIR-05TP3 Washington, DC 20361-3300	1
CDR J.M. Owens Naval Research Lab Code 5510 4555 Overlook Avenue, SW Washington, DC 20375-5000	1
Dr. Raja Parasuraman Dept. of Psychology Catholic University of America Washington, DC 20064	1
Dr. John Reising Cockpit Integration Directorate Code WRDC/KTC Wright-Patterson AFB, OH 45433	1
Dr. Joel Warm University of Cincinnati Dept. Of Psychology Cincinnati, OH 45221	1
Naval Air Warfare Center - Aircraft Division (2 for Code 8131)	2

DISTRIBUTION LIST (Continued)

	No. of Copies
Defense Technical Information Center ATTN: DTIC-FDAB Cameron Station BG5 Alexandria, VA 22304-6145	2
Center of Naval Analysis 4401 Fort Avenue P.O. Box 16268 Alexandria, VA 22302-0268	1
1299th Physiological Training Flight Malcolm Grow USAF Medical Center Andrews AFB, Washington, DC 20331-5300	1